

Tartu Ülikool

Maailma keelte ja kultuuride kolledž

Martin Mala

Manuaali

„KoGloss:

korpuspõhine kollaboratiivne konstruktsioonide glossaarium

ülikoolide keeleõppes ja kutsetegevuses“

tõlge ja tõlkeanalüüs masintõlke seisukohast

Magistritöö

Juhendaja: Terje Loogus, PhD

Tartu 2017

SISUKORD

Sissejuhatus	3
1. Sihttekst.....	4
2. Magistritöö teoreetilis-metodoloogilised lähtekohad.....	48
3. Masintõlkest	50
3.1. Masintõlke liigid.....	50
3.1.1. Reeglipõhine masintõlge.....	50
3.1.2. Statistiline masintõlge.....	52
3.1.2.1. Google'i tõlge.....	53
3.1.2.2. Bingi tõlge.....	55
3.1.3. Tehisnärvivõrgu põhine masintõlge	56
3.2. Masintõlke võimalused ja kasulikkus tõlkimisel.....	57
4. Tõlkeanalüüs	60
4.1. BLEU.....	60
4.2. Masintõlke võimekus keskse terminoloogia tõlkimisel	61
4.3. Masintõlke võimekus KoGlossi manuaali tõlkimisel.....	63
4.4. Masintõlketööriistade kasulikkus nelja näite põhjal	66
4.4.1. Näide 1.....	66
4.4.2. Näide 2.....	67
4.4.3. Näide 3.....	69
4.4.4. Näide 4.....	71
4.5. Üldised tõlkeprobleemid	74
4.6. Järeldused	78
Kokkuvõte	80
Kasutatud allikad.....	82
Resümee	85

Sissejuhatus

Käesoleva magistritöö eesmärgiks on tarbeteksti tõlke näitel välja selgitada, kas ja millisel määral on töö koostamise ajal (aastatel 2016 kuni 2017) kättesaadav masintõlge abiks tõlkimisel saksa-eesti suunal. Selle eesmärgi elluviimiseks on tõlgitud saksa keelest eesti keelde „KoGloss, korpuspõhine kollaboratiivne konstruktsioonide glossarium ülikoolide keeleõppes ja kutsetegevuses“ manuaal. Tegu on tekstiga, mida võiks lugeda „keskmisest keerulisemaks“. Seejuures ei eeldata, et masintõlge oleks suuteline andma kohe ideaalset tõlget. Analüüsimisel keskendutakse eelkõige sellele, kas masintõlge aitab selle laadse teksti tõlkimisel aega ja vaeva kokku hoida või mitte. Idee KoGlossi manuaali tõlkimise jaoks andis minu juhendaja Terje Loogus, kes on samuti selle projekti liige. Minul endal oli soov teha midagi praktilist, mis oleks seotud minu elukutse valikuga. Minu päris algne plaan oli tõlkida ühte saksakeelset romaani, kuid see ei läinud kokku tõlkeõpetuse magistritöö nõuetega. Seega langes minu valik käesoleva teema kasuks.

Magistritöö sissejuhatusel järgneb KoGlossi saksakeelse manuaali tõlge eesti keelde. Teises peatükis tutvustatakse magistritöö teoreetilis-metodoloogilis lähtekohti. Kolmandas peatükis tutvustatakse magistritöö kirjutamise ajal olemasolevaid masintõlke liike ja kahte kõige populaarsemat masintõlketööriista. Peatüki lõpus käsitletakse masintõlke võimalusi ja võimekust tõlkimisel erinevatest vaatepunktidest ning selle lõpus räägitakse ka natukene minu enda kogemustest masintõlkega. Sellele järgneb tõlkeanalüüsi peatükk, kus esimeses alapeatükis tutvustatakse masintõlke hindamise meetodit BLEU ning teises alapeatükis tutvustatakse töö keskset terminoloogiat ja kuivõrd võimekas on masintõlge selle tõlkimisel. Kolmandas käsitletakse masintõlke võimekust ja kasulikkust KoGlossi manuaali tõlkimisel ning üldiseid masintõlkega seonduvaid probleeme, tähelepanekuid ja erinevaid iseärasusi. Neljandas alapeatükis vaadatakse masintõlketööriistade võimekust nelja näite põhjal. Sellele järgnevas alapeatükis käsitletakse KoGlossi manuaali tõlkimise käigus tekkinud üldiseid tõlkeprobleeme ja küsimusi. Sellele järgneb järelduste osa. Lõpuks magistritöö kokkuvõte, kus antakse ülevaade töö käigus saadud tulemustest. Sellele järgnevad kasutatud allikad ja töö saksakeelne resümee.

1. Sihttekst

I. MIS ON KOGLOSS? ÜLDINE METOODIKA

Oskuskeelte valdkonnas lasub põhirõhk endiselt üksikute väljendite kogumisel, tõlkimisel ja seletamisel. Erinevamate kutsetegevuste terminoloogiate kohta leidub palju teoseid. Aga kuidas on lood, kui minna üksikust sõnast kaugemale? Kelle poole tuleks pöörduda, kui peaks tekkima küsimusi, mis ei ole seotud üksikute väljenditega, vaid sõnade kombinatsioonide ja keeleliste mustritega? Nõndanimetatud „mitte-spetsialistil“ võib olla raske otsustada, milliseid struktuure peaks kasutama saksa majanduskeeles, kui on vaja rääkida protsentidest, inflatsioonist ja tööhõive näitajatest. Üllatada võib, et saksa keeles on võimalikud konstruktsioonid *Liquidität bereitstellen*, *zuführen* ja *jemanden damit versorgen*. Ja mida tähendab see, kui majanduskasv on negatiivne? Tavapärased allikad pakuvad sellistele küsimustele ainult piiratud vastuseid. Sama probleem tekib ka keeleõppes. Õpetamisel ja õppimisel on hakatud isoleeritud sõnavara tuupimisest eemalduma ning pöörduma koos esinevate sõnade ja keeleliste konstruktsioonide vahenduse poole. Kuid sellise süstematiseeritud juurdepääsu loomine on valdkond, milles on veel palju kasutamata potentsiaali.

KoGloss kui keele õppemeetod pakub võimalusi nende vajaduste rahuldamiseks.

Mis on KoGloss?

KoGlossi nimes on ühelt poolt esindatud glossaarium kui leksikograafiline žanr, teisalt tähistab element „Ko“ projekti kolme põhiaspekti: See tähistab nii kollaboratiivsust, korpust kui ka konstruktsioone, ning viitab mitme grupi koostööle (kollaboratiivne), mis tegelevad keeleüksustega minnes üksikust sõnast kaugemale (konstruktsioonid), kusjuures analüüsi aluseks on digitaalsed tekstikogud (korpus).

KoGloss töötab madalalävelise digitaalse tekstianalüüsi vormis. Kõigepealt kogutakse kokku materjali, mis esindab ja hõlmab vastavat eriala piisaval määral. Kollektsoon võib koosneda näiteks ajakirjandusest pärit erialalistest tekstidest, erialastest väljaannetest ja ametlikest väljaannetest, mis on internetis kättesaadavad või tänu oma isiklikule taustale saadud. Seejärel analüüsitakse materjali vabalt kättesaadava

keeletarkvara abil, et teha kindlaks eriliste keeleliste mustrite esinemine ja funktsioon. Tulemused talletatakse digitaalses glossaariumis.

KoGlossi meetodi eripäraks on see, et eri valdkondadest ja erialadelt pärit autentse kirjaliku materjali keelelist eripära saab uurida ja talletada. Vaadeldava eriala tavapärased keelelised mustrid saab tuvastada ja teha kättesaadavaks ilma, et see piirduks üksikute erialaste terminite tasandiga. Järgmiseks eeliseks on see, et tekstide kogumine ja analüüs ei vaja erilisi keeleteaduslikke teadmisi, pärast lühikest koolituse perioodi saavad eri teadmiste tasemega kasutajad KoGlossi rakendada.

KoGlossi meetodi potentsiaalsed kasutajad on need, kes peavad tööalaselt ja/või täiendõppes pöörama erilist tähelepanu teksti indekseerimisele ja koostamisele. Selleks on keeleõppijad, keeleõpetajad ja spetsialistid nagu tõlkijad, tõlgid, ajakirjanikud, suhtekorraldajad, sekretärid, võõrkeele korrespondendid, terminoloogid. KoGloss aitab nii emakeelsetel kõnelejatel kui ka võõrkeele õppijatel saavutada kõrgemat keelelist teadlikkust ja olla toeks glossaariumi kujul väga konkreetsete küsimuste puhul. Meetod võimaldab luua kiiresti ja sihipäraselt uusi eralakeelseid domeene oma või võõrkeeles. Seega võib eeldada, et oma ja võõrkeelte analüüsimise võime suureneb oluliselt ja on otseselt kasulik kommunikatsiooniga seotud erialadel.

Käesolevas manuaalis tutvustatakse samm-sammult keele õpetamise meetodit KoGloss, mis koosneb kolmest põhietapist: korpuse koostamine, korpuse analüüs ja keeleliste konstruktsioonide kirjeldamine glossaariumis.

Lisateavet saab järgmisel aadressil:

<http://www.uni-due.de/kogloss.eu/index.php>

II. KORPUSE KOOSTAMINE

Selleks et kasutada KoGlossi spetsiifiliste keeleliste mustrite määramiseks on esmalt vaja luua digitaalne tekstide kollektsioon (korpus). Hariduses ja kutsetegevuses saab seda teha kahel viisil: tekstikorpuse koostavad õpetajad või projektijuhid ja esitatakse õppijatele või projekti kaastöötajatele valmistootena, või tekstikorpus koostatakse koostöös õppijate või projekti kaastöötajatega.

Enne korpuse koostamisega alustamist, on oluline kõigepealt määratleda tekstide valiku ja andmeallikate kindlad kriteeriumid. See puudutab eelkõige tekstide sisu, juurdepääsu nendele tekstidele ja korpuse mahtu (nt tekstide arv).

1. TEKSTIDE VALIK

Tekste saab hankida internetist või skaneerida. Tuleb määrata kindlaks, millistele kriteeriumidele korpusesse lisatavad tekstid vastama peavad (näiteks teema, teksti liik, meedium, teksti pikkus, teksti koostamise aeg jne).

Tagamaks korpuse temaatilise ühtsuse saab luua interneti otsingusõnade nimekirja. Tuleb siiski märkida, et tekste, milles otsingusõna esineb ainult juhuslikult, ei lisata korpusesse.

Samuti on oluline vältida selliste tekstide lisamist, mis põhinevad samal konkreetsel tekstiliigil ja allikal ning, mis keeleliselt ei varieeru vähe või üldse. Nii saab vältida, et mõned konstruktsioonid esinevad korpuses kõrge sagedusega, ilma et see viitaks laialdasele kasutusele.

Tehnoloogilistel põhjustel ei ole soovitatav lisada sellist PDF-faili, mille tekst on „pildistatud“ või reavahe vormingu tõttu on konverteerimise protsess liiga mahukas.

2. KONVERTEERIMINE, ARHIVEERIMINE JA TEKSTIDE REDIGEERIMINE

Tekstide konverteerimisel tuleb järgida järgmist mustrit: PDF → WORD → TXT.

TXT-failid salvestatakse süstematiseeritud failinimedega. Selleks on kõigepealt vaja luua eraldi dokument, mis sisaldab kasutatud allikate **lühendeid**, näiteks:

Allikas	Lühend
SPIEGEL	SP
Bundesministerium der Finanzen	BMF
Financial Times	FT

Failinimi võib olla üles ehitatult järgmiselt:

Allikas(lühend)_aasta_kuu_avaldamise kuupäev_pealkiri(lühend)

Näiteks: SP_2008_05_15_Konjunktur.txt

Korpusesse lisatava teksti **redigeerimiseks** on soovitatav WORD-formaat, sest see sobib paremini kui TXT-formaat. Dokument peab lõpuks sisaldama ainult lihtteksti. Selleks on vaja:

- Ühendada üksikud sõnade fragmendid, mis tekivad poolitamisel rea lõpus,
- Kustutada järgmised teksti elemendid: lehekülje numbrid, päised ja jalused, ääremärkused, autorite nimed, kirjanduse andmed, foorumi või kasutajate kommentaarid, lingid, tabelid ja pildid, lõigunumbrid, võõrkeelsed lõigud, erisümbolid (näiteks täpploetelud) ja matemaatilised valemid. Lisaks veel tuleb kustutada teksti ja lõigu pealkirjad, piltide seletused ja täpploetelud, välja arvatud juhul, kui need esinevad tervete lausete kujul.
- Lisada punktid teksti ja lõikude pealkirjadele, piltide seletustele ja nimekirjadele, mis esinevad tervete lausetena ning integreerida need teksti.

Enne redigeerimist saab lisada noolsulgudega allika järgmisel kujul:

<Salvestatud failinimi: Täielik link>

Näiteks:

<SP_2008_05_15_Konjunktur:<http://www.spiegel.de/wirtschaft/0,1518,553373,00.html>
>

Pärast redigeerimist võib dokumenti konverteerida TXT-formaati. Oluline on pöörata tähelepanu ühtsele kodeerimissüsteemile **UTF-8**.

Seejärel saab sisestada TXT-faili eraldi loodud allikate loendisse järgmisel kujul:

Failinimi. Täispikk nimi. Täielik link (Päringu kuupäev)

Näiteks:

SP_2008_05_15_Konjunktur. *Deutsche Wirtschaft wächst überraschend stark.*
<http://www.spiegel.de/wirtschaft/0,1518,553373,00.html> (22.01.2011)

KoGlossi meetodi järgmine samm on loodud tekstikorpuse uurimine keeletarkvaraga *AntConc*, mida käsitletakse järgmises peatükis.

III. KORPUSE ANALÜÜSIMINE ANTCONCIGA

1. MIS ON ANTCONC?

Eriliste keeleliste mustrite kindlakstegemiseks uuritakse KoGlossi meetodi järgi ise koostatud digitaalseid tekstikogusid (vt ptk II korpuse loomine). Selleks et nn digitaalseid korpuseid oleks võimalik kasutada, on olemas erinevaid analüüsiprogramme. Nende eesmärk on koguda teavet teatud keeleliste mustrite kohta varem kogutud tekstidest. Päringuid saab sisestada üksikute sõnade, sõnade osade või sõnajärgede vormis. Sellised analüüsitarkvarad on erineva keerukusega ja pakuvad erinevaid analüüsivahendeid. Tasuta ligipääs on näiteks programmidele *Simple Concordance Program* (SCP) või *KwiCFinder*, samas programmid nagu *ParaConc* või *Wordsmith Tools* on tasulised. Sõltuvalt vajadustest saab otsustada, milline programm sobib kõige paremini. KoGloss kasutab tavakasutajalegi vabalt kättesaadavat keeletarkvara *AntConc*.

AntConci töötas välja *CELESE* (Center for English Language Education in Science and Engineering) juht professor Laurence Anthony Waseda Ülikoolis Tokyos ja on sobilik eelkõige väiksemate tekstikogude (kuni 20 miljonit sõna) jaoks. See on laialdaselt kasutatav programm digitaalsete tekstikogude analüüsimisel, mida on lihtne kasutada, ei ole vaja arvutisse installeerida ning saadaval on üksikasjalikud juhised ja muud täiendavad ressurssid (vt.

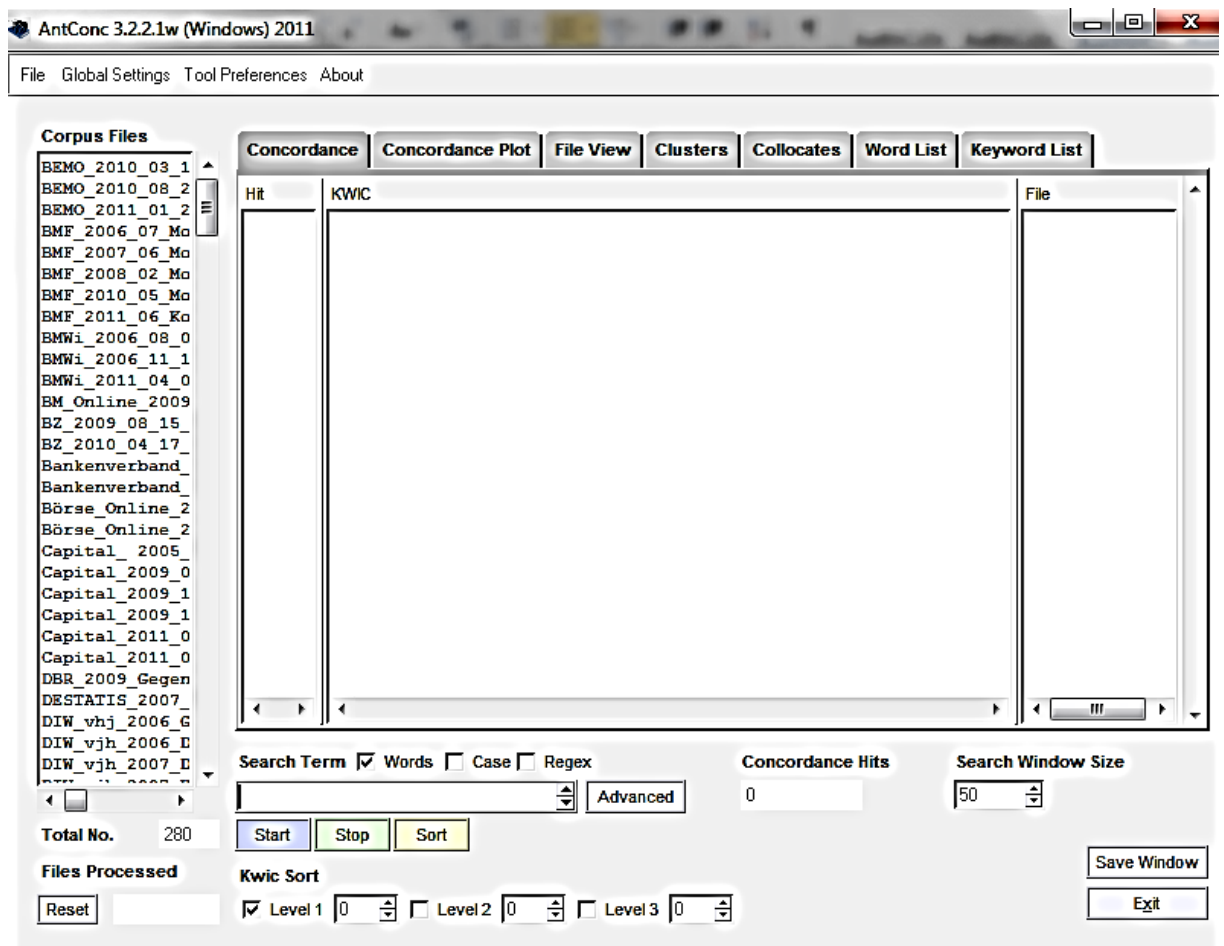
http://www.antlab.sci.waseda.ac.jp/software/README_antconc3.1_german.pdf).

Pakutavad tööriistad võimaldavad luua sõnade nimekirju, klastreid, konkordantse, kollokatsioone ja anda üksikasjaliku ülevaate otsingutulemuste kontekstidest.

2. ANTCONCI INSTALLEERIMINE JA KÄIVITAMINE

Tarkvara saab tasuta alla laadida *AntConci* autori koduleheküljelt (<http://www.antlab.sci.waseda.ac.jp/software.html>). Pärast allalaadimist saate programmi käivitada (paigalduseta) klikkides .exe-failile. Selle struktuur põhineb tuttavalt *Windowsi* paigutusel. Programm kasutab ka mitmeid erinevaid *Windowsi* tuttavaid otseteid.

Pärast avamist ilmub selline aken:



Corpus Files (vasakus servas): Kuvatakse ülevaade laetud tekstifailidest, mille koguarvu näidatakse all *Total No.* kõrval.

Kaardid: Kaartidena on loetletud erinevad tööriistad, mille abil saab näiteks luua konkordantse, määrata kindlaks sõna rühmad, leida koosinemist või sagedust. Pärast otsingut ilmuvad tulemused selle akna keskmisesse tulpa.

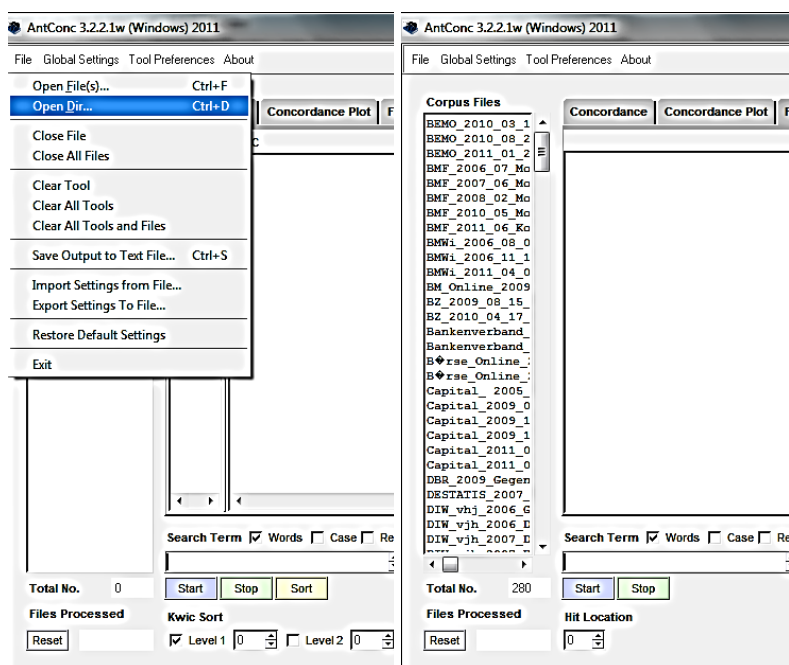
Päringukast ja otsingu seaded (all keskel): Sellesse saate sisestada päringuid ja neid otsima hakata. Igale tööriistale saab määrata erinevaid otsingu tingimusi. Üksikasjad on selgitatud peatükis *Otsingu sisestamine*.

3. ENNE ANALÜÜSI ALGUST

a) Tekstikogu sisestamine

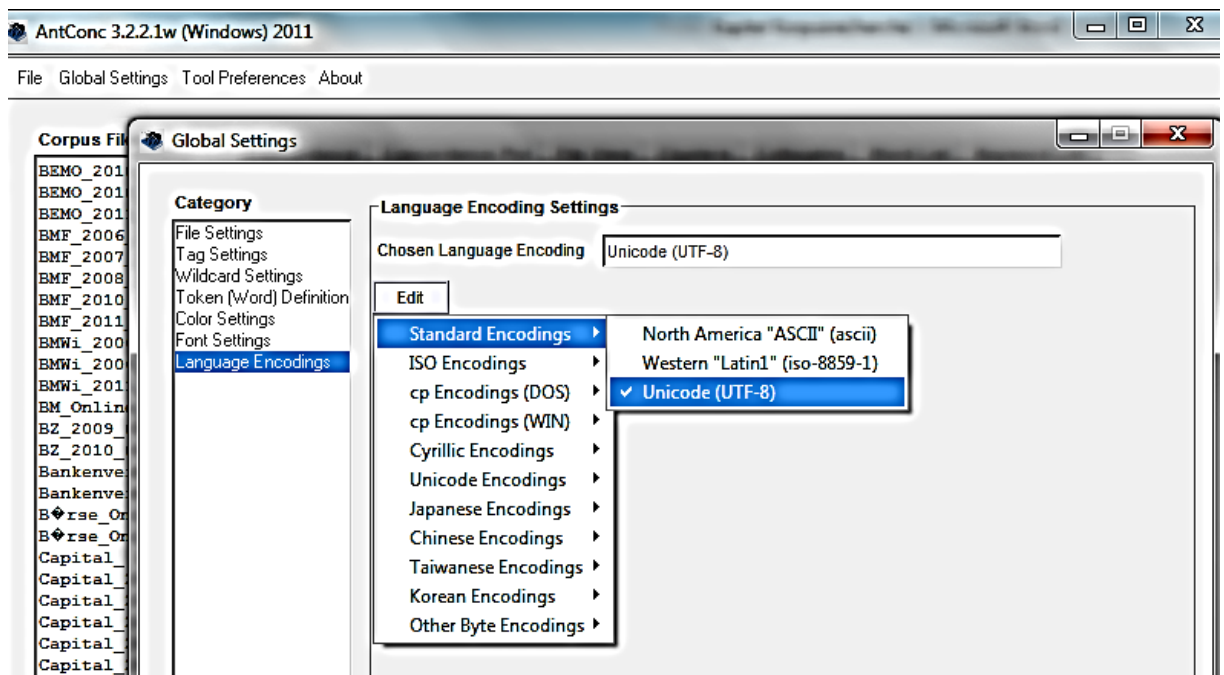
Enne kui *AntConc*iga analüüsi alustada saab, tuleb esmalt laadida tekstid programmi. Menüüvaliku **File** all saab *Open File(s)* ja *Open Dir...* valida ning laadida arvutisse salvestatud korpuse faile või terveid korpuse failie sisaldavaid kaustu.

Laetud tekstifailide nimekiri ilmub vasakule.

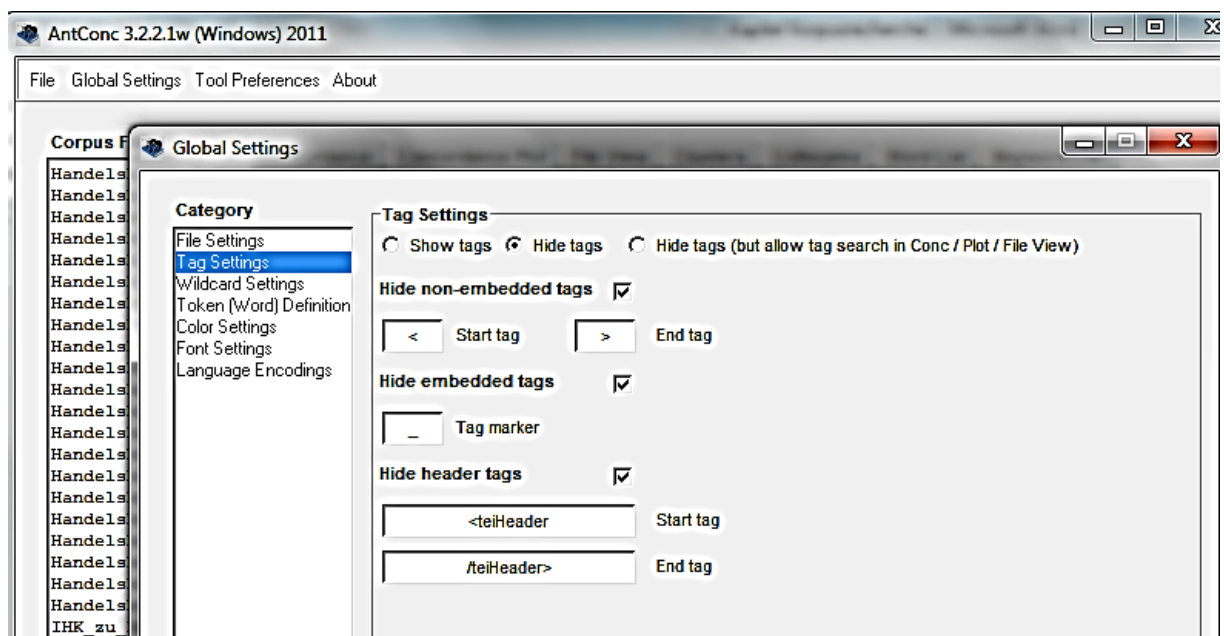


b) Seaded

Menüü valiku **Global Settings** all saate teha seadistusi, et tagada andmete korrektset töötlemist *AntConc*is. Nendeks on näiteks *Language Encoding* või *Tag Setting* (allpool). Kogu info sätete kohta leiate *AntConc*i käsiraamatust (peatükk „Menüü valikud“).



Language Encodings: Tähemärkide korrektse kuvamise tagamiseks valige kodeering UTF8.



Tag Settings: Seade *Hide Tags* abil saate märkida noolsulgudega, millised teksti osad jäävad otsingul välja, näiteks meta-andmed, nagu <autor>, <pealkiri> jne.

c) Tööriistad ja nende funktsioonid

KoGloss meetodi põhiülesannete täitmiseks kasutatakse tööriistu *Word List*, *Concordance*, *Clusters*, *Collocates* ja *File View*. Neid saab kasutada klõpsates vastava kaardi peale.

Tööriist *Word List* võimaldab koostada korrapärastatud nimekirja tekstifailides esinevatest sõnadest. Sõnu saab sorteerida sageduse või tähestikulises järjekorras, kas sõna alguse või lõpu järgi ja järjekorda saab ümber pöörata.

Tööriista *Concordance* kasutatakse nn KWIC-ridade loomiseks. Seejuures loetletakse read, milles võtmesõna kontekstis (KWIC: **key word in context**) esineb ja annab teavet, millisel viisil sõna või keelestruktuuri tekstikogus kasutatakse.

Funktsioon *Clusters* võimaldab luua korrapärastatud loetelu sõnade rühmadest, mis esinevad valitud tekstifailides otsingusõna ümber. Sub-funktsiooniga *N-Grams* on võimalik, ilma ühelegi kindlale sõnale või mustrile keskendumata, tuvastada tihti koos esinevaid sõnade rühmi. Võimalus on valida sõnade rühmade minimaalne ja maksimaalne pikkus (sõnade arv) ja kuvatavate sõnade rühmade minimaalne sagedus.

Tööriist *Collocates* võimaldab luua nimekirja, kus on näha otsingusõnaga sageli koos esinevaid sõnu või mustreid. Seejuures on võimalik määrata nende pikkust, mida otsingul arvestatakse.

Funktsioon *File View* võimaldab igal ajal leida otsingusõna juurde kuuluvat teksti. Ilmub terviklik vaade vastavast tekstist, mis annab ülevaate esiletõstetud otsingusõna tervest keelelisest ümbrusest.

4. ANTCONCI KASUTAMINE: TOIMIMINE JA FUNKTSIOONID

4.1. PÄRINGUTE ESITAMINE

a) Päringud otsingusõnaga ja ilma

Päringuid alustatakse üldiselt klõpsuga nupule **Start**. Päringuid on võimalik esitada nii otsingusõnaga kui ka ilma. Sõltumata otsingusõnast saab luua nimekirju tekstikogus sageli esinevatest sõnadest (**Word List**). Peale selle on võimalik teha kindlaks sagedased sõnade rühmad, selleks tuleb kaarti **Clusters** all käivitada päring funktsiooni **N-Grams** kaudu.

Kui soovitakse leida teavet konkreetse otsingusõna kohta, siis tuleb see sõna sisestada sisendreaale ja alustada päringuga klõpsates nupule **Start**. Funktsiooniga **Collocates** võimaldab *AntConc* tuvastada otsingusõnaga sageli esinevaid sõnu või otsida funktsiooni **Cluster** kaudu sõnarühmi, kus otsingusõna tihti esineb. Juhul kui kuvatakse kõik read, milles otsingusõna esineb, siis alustatakse päringut kaartis **Concordance**.

Selleks et saada rohkem teada otsingusõnade ja/või sõnarühmade kontekstist, tuleb neile peale klõpsata ja kaardid **Concordance** ning **FileView** annavad ülevaate nende keelelisest keskkonnast või tervest tekstist.

Täpsem teave funktsioonide kohta järgneb allpool.

b) **Wildcardside kasutamine**

Otsingusõnadega päringute puhul on võimalik kasutada nn **Wildcardse**, millel on kohatäitja funktsioon. **Wildcardside** abil on võimalik arvestada sama sõna erinevaid käände- või pöördelõppe.

Võimalik on täpselt määratleda, kui palju märke kohatäitjad katavad ja kus need võivad esineda.

Ülevaade *AntConc*is automaatselt lisatavatest kohatäitja funktsioonidest:

+ üks või mitte ühtegi tähemärki;

* üks või rohkem tähemärki;

? täpselt üks märk;

@ üks või mitte ühtegi sõna;

täpselt üks sõna;

| otsingusõna 1 VÕI otsingusõna 2.

Lisades otsingupäringusse kohatäite „+“, arvestatakse kõiki sõnu, millel on selles kohas täpsustamata või konkreetsed täiendavad tähemärgid. Seega leitakse päringuga *koha+* sõnavormid *kohal*, *kohas* ja *kohad*. Juhul kui kohatäide jääb tühjaks, kasutatakse sisendit „?“.

Lisades otsingupäringusse kohatäite „*“, arvestatakse kõiki sõnu, millel on selles kohas üks või enam täpsustamata tähemärki. Seega laieneb nimekiri ka sellistele sõnadele nagu *kohale*, *kohased* ja *kohakuti* jms.

Kohatäitjaid „@“, „#“ ja „|“ kasutatakse tervete otsingusõnade või otsingusõnade kombinatsioonide puhul. Ühe või mitte ühegi sõna arvestamiseks otsingusõna läheduses kasutatakse kohatäidet „@“, samas kui täpselt üks sõna määratakse „#“-ga.

Kui on soov teha päring kahest tuntud otsingusõna kombinatsioonist, kasutatakse kohatäidet „|“, mis tähendab loogiliselt „või“. Leitakse need kohad kus üks või teine otsingusõna esineb, kusjuures nende juhtude puhul, kus mõlemad sõnad esinevad, puudub eriline märgistus. Nii saadakse näiteks päringuga *küsimus|vastus* kõik tulemused, kus esineb *küsimus* ja kõik tulemused kus esineb *vastus*.

c) Abi erandite puhul

Sihipäraste ja tõhusate päringute tegemiseks, tuleb arvestada teatud keeleliste eripäradega.

See käib eriti tegusõnade, nimisõnade ja omadussõnade muutelõpu kohta. Selleks et ilmuksid kõik võimalikud sõna põhivormi variandid, peavad sõna võimalikud eri lõpud olema asendatud kohatäidetega. See näeb välja üksikasjalikult järgmiselt:

Nimisõna näide:

Sisend: *System++*

Leitud sõnavormid: *System, Systeme, Systems, Systemen*

Näidisread¹: [...] *koexistierten die **Systeme** friedlich* [...]

*Das regulative **System** der Versicherungswirtschaft* [...]

*Die Hauptkomponenten des **Systems** bilden* [...]

Päring näitab tulemusi, milles on arvesse võetud sõna nii ainsuses kui ka mitmuses ja erinevates käänetes. Kui soovitakse leida uusi sõna koosluseid, milles otsingusõna on tüvi, tuleb valida kohatäide, mis lubab erinevaid märke. Seega annab päring *System** tulemused nagu *systematisch, Systematisierung, systemerelevant* jms.

Omadussõna näide:

Kohatäidet ,+‘ pannakse omadussõnade puhul sageli üksteise järele, sest omadussõnade lõpud võivad olla pikemad.

Sisend: *klein+++++*

Leitud sõnavormid: *klein, kleines, kleinen, kleiner, kleineren, kleinsten*, jms

Näidisread: *Wenn nur ein **kleiner** Teil der 2,3 Millionen* [...]

*Dies dürfte vor allem **kleinere** Börsen in Europa belasten* [...]

[...] *sei die Abweichung **kleiner** als 0,5 % gewesen* [...]

¹ Kõik näited pärinevad KoGlossi majanduse saksa keele korpusest. *AntConci* kasutamisel tuleb keele tunnuseid.

Tegusõna näide:

Sisend: ++*such*+++

Leitud sõnavormid: *suchen, sucht, suchte, suchten, gesucht*

Näidisread: *Jede Regierung **sucht** nun nach der geeigneten Strategie [...]*

*Deshalb wird 2009 nach einem Maß **gesucht** [...]*

*[...] wer einen **sucht**, rechnet sich mehr Chancen aus [...]*

Antud otsingu tulemuste saamisel on arvestatud nii sõna ainsuse kui ka mitmuse, eri ajavorme ja eri isikulisi vorme. Verbide puhul on eriti oluline katta kohatäitjatega ka võimalikud minevikuvormid (näiteks mineviku kesksõnale *gesucht* eelneb topelt ,+‘).

Kui tahetakse leida uusi sõnade koosluseid, milles verb on tüvi, tuleb valida kohatäide, mis lubab igasuguseid märke. Sisend **such** tulemusteks on näiteks *untersuchen, Untersuchung, Arbeitssuchende, heimgesucht* jne. Paigutades kohatäite ,*‘ tüve ette arvestatakse ka eesliiteid.

Konkreetne juhtum algustähega

Üldiselt tuleb märkida, et muutuva sõna tüve puhul (nt *Plan: Pläne*) ei ole sel viisil võimalik katta kõiki vorme ja võib-olla tuleb sooritada eraldi otsinguid. Nagu näiteks *gehen*, mille vorme *gegangen* või *ging* ei saa verbi tüve järgi leida. Saksa keeles on olemas ka lahutatava eesliitega verbid (nt *einbrechen* või *losfahren*), mis teatud kontekstis paiknevad lauses kahes erinevas kohas, näiteks minevikuvormis: „*Die Erzeugung im Bauhauptgewerbe **brach** im März um 10,8 % **ein**.*“ (Handelsblatt Online, 09/05/2005)

Tuleb arvestada ka muid üksikute keelte erivorme.

AntConc võimaldab koostada nn *Lemma*-nimekirju, mis katavad kõiki sõna võimalikke vorme ja hõlbustada nii järgnevat otsingut.

4.2. WORD LIST

Selleks et koostada nimekiri kõikidest tekstikogus esinevatest sõnavormidest tuleb avada kaart **Word List** ja alustada päring nupuga **Start**. Seejuures ei sisestata otsingusõna. Otsingu jaoks võite teha järgmisi seadistusi:

- **Sort by:** Sõnu saab sorteerida tähestikuliselt või sageduse järgi. Sorteerimise suunda on võimalik ümber pöörata (*Invert Order*).
- **Treat all Data as Lowercase:** Kui see seadistus on aktiivne, ei tehta vahet suur- ja väiketähedel.

Tulemuseks on järgmine teave:

AntConc 3.2.2.1w (Windows) 2011

File Global Settings Tool Preferences About

Corpus Files

- SPIEGEL_Online
- SPIEGEL_Online
- SPIEGEL_Online
- SPIEGEL_Online
- SPIEGEL_Online
- SPIEGEL_Online
- SPIEGEL_Online
- SPIEGEL_Online
- SPIEGEL_Online
- SPIEGEL_Online
- STERN_2005_06_
- STERN_2005_06_
- STERN_2005_06_
- STERN_2005_08_
- STERN_2006_12_
- STERN_2007_01_
- STERN_2007_09_
- STERN_2011_06_
- SUITE101_2010_
- Tagesschau_201
- Tagesschau_201
- WAMS_2008_05_1
- WELT_Online_20
- WELT_Online_20
- WELT_Online_20
- WELT_Online_20
- WIWO_2011_06_2

Concordance Concordance Plot File View Clusters Collocates Word List Keyword List

Hits Total No. of Word Types: 35553 Total No. of Word Tokens: 401347

Rank	Freq	Word	Lemma Word Form(s)
1	15522	der	
2	15202	die	
3	7906	und	
4	6640	in	
5	5351	den	
6	4561	im	
7	4549	von	
8	3737	Die	
9	3719	für	
10	3626	zu	
11	3610	auf	
12	3443	des	
13	3418	sich	
14	3087	das	
15	2813	ist	
16	2675	mit	
17	2343	nicht	

Search Term ☒ Words ☐ Case ☐ Regex

Advanced

Display Options

☐ Treat all data as lowercase

Total No. 280

Files Processed

Reset

Hit Location

Search Only 0

Sort by

Sort by Freq

☐ Invert Order

Save Window

Exit

Kast **Word** näitab teksti kõiki sõnavorme.

Rank: näitab iga sõnavormi pingerida seoses nende sagedusega tekstikogus.

Freq: näitab otsitava sõnavormi esinemise sagedust tekstikogus.

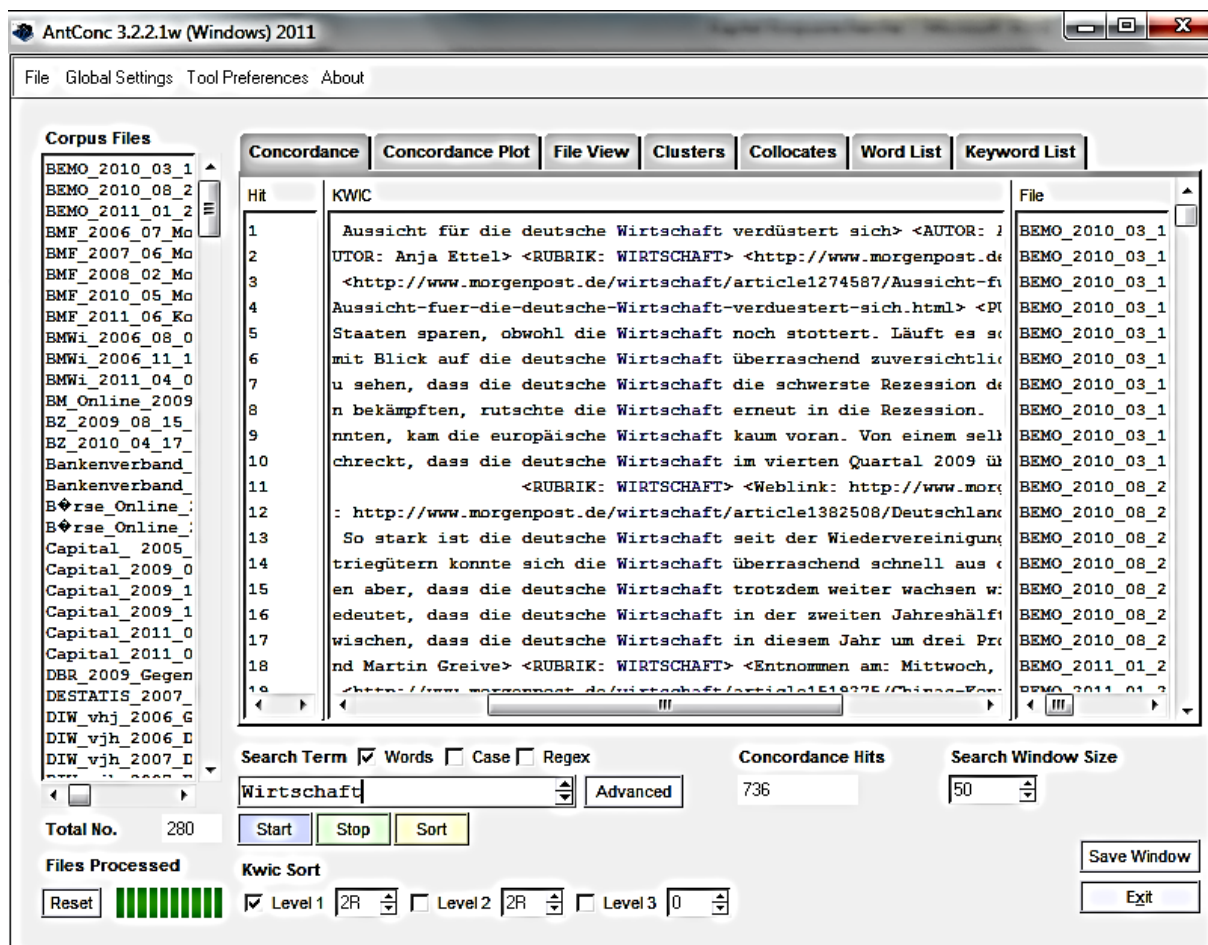
Types: tekstides sisalduvate sõnavormide koguarv.

Tokens: tekstides sisalduvate sõnade koguarv.

4.3. CONCORDANCE

KWIC-ridade nimekirja saamiseks, avage kaart **Concordance**. Päringut saab alustada võtmesõna sisestamisega. *Concordance Hits* näitab leitud konkordantside arvu.

Otsingu seadeid saab kohandada järgmiselt:



Case: Kui see seade on valitud, arvestatakse otsingusõna suur- ja väiketähti, mida muidu ignoreeritakse.

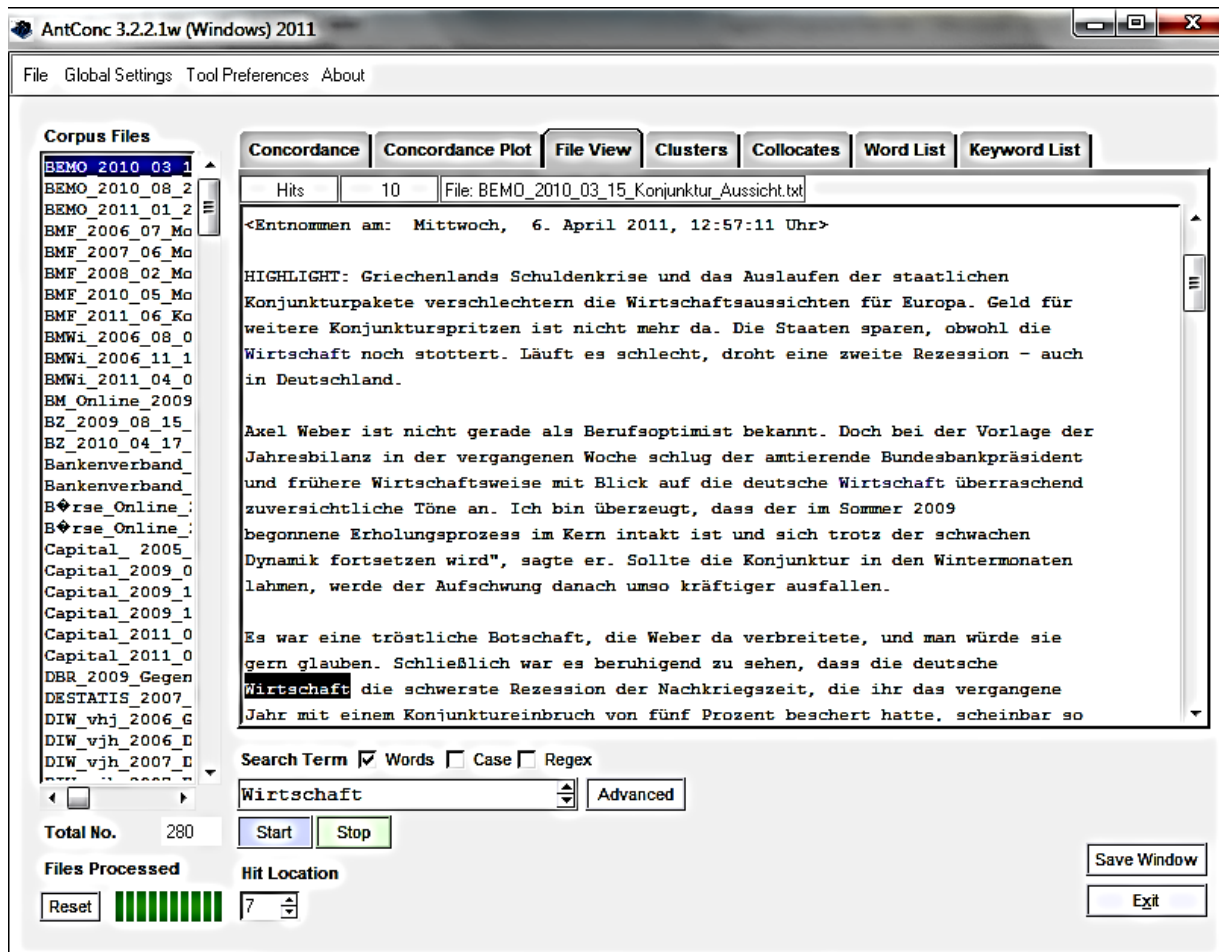
Words: Otsitav väljend on terve sõna. Kui see pole valitud, võib leida sõna fragmente.

Kwic Sort: Saab valida kui palju sõnu otsingusõnast vasakule/paremale esile tõstetakse.

Selleks et kuvada ennem tehtud otsingu tulemusi, tuleb klõpsata *Cluster-*, *Word-List-* või *Collocates*-funktsioonide peale ning seejärel ilmub automaatselt kaart *Concordance* koos vastavate KWIC-riadega.

4.4. FILE-VIEW

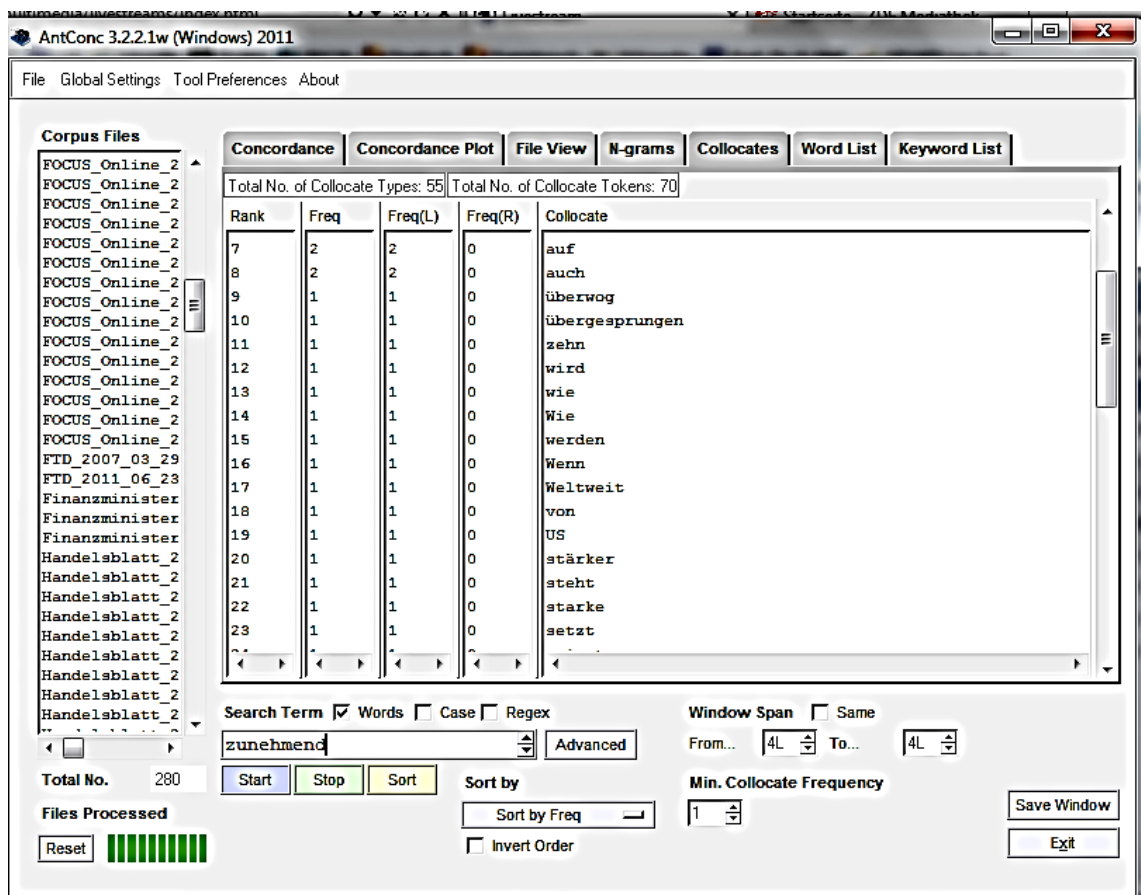
Selleks et vaadata otsingusõna keskkonda algtekstis, tuleb klõpsata vastava sõna või mustri peale ja vaade läheb üle **File-View** peale. Aknas kuvatakse nüüd kogu tekstifaili, milles vastav tulemus on leitud.



4.5. COLLOCATES

Kaarti *Collocates* all on võimalik luua nimekirjad sõnadest, mis teatud otsingusõnaga koos kõige sagedamini esinevad.

Seejuures Saate teha järgmisi seadistusi otsimiseks ja sorteerimiseks:



Window Span all saab valida otsingusõnast paremal ja vascul asuvate sõnade arvu (nt *From... 4L To...4L*=pikkus: neli sõna).

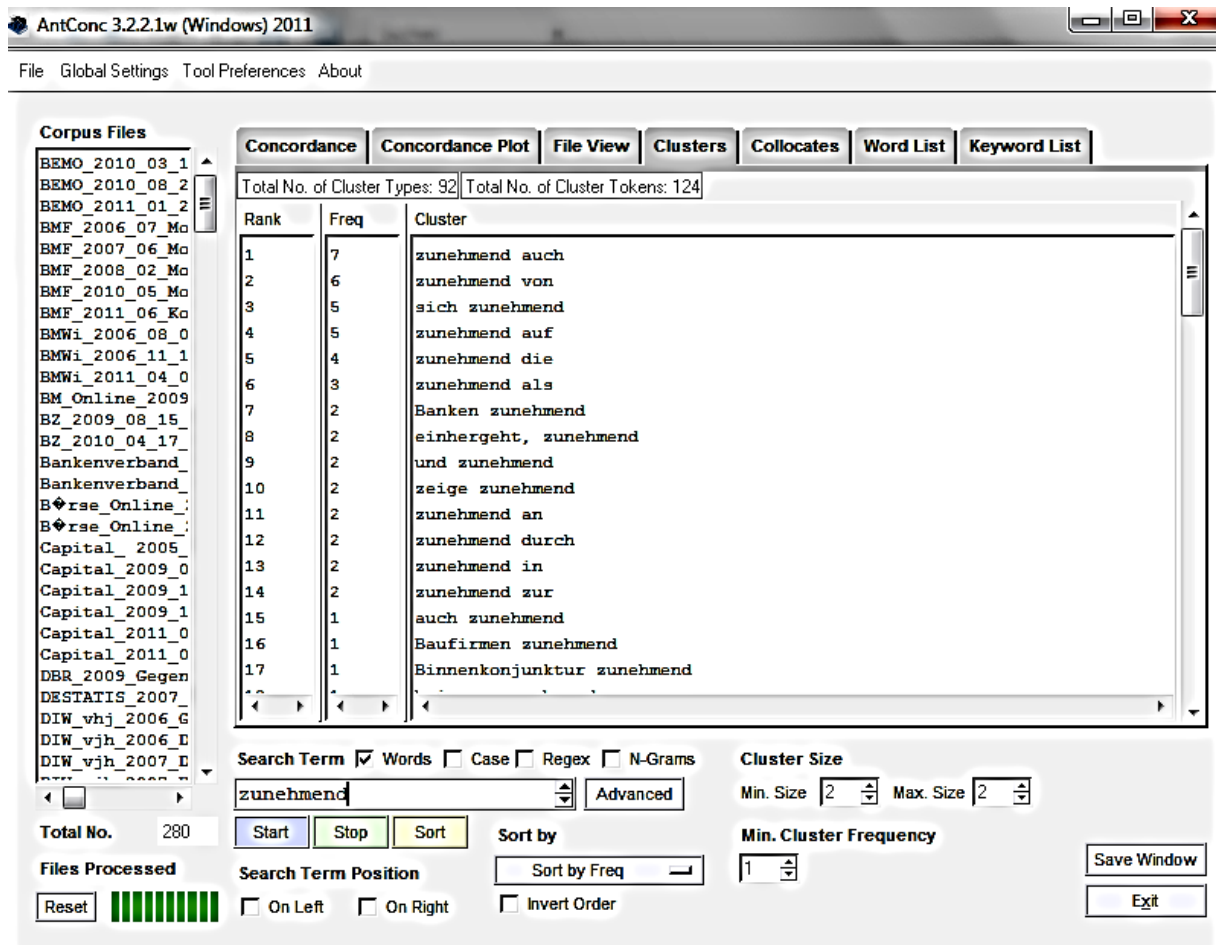
Selleks et hoida minimaalne ja maksimaalne pikkus sama, klõpsake ruutu *same*.

Min. Collocate Frequency määrab partnerite minimaalse koosinemise sageduse.

4.6. CLUSTERS/N-GRAMS

a) Päring otsingusõnaga

Funktsioon **Clusters** koostab loetelu sõnarühmadest, mis esinevad valitud tekstifailides ühe otsingusõna ümber. Selleks sisestatakse vastav sõna otsingukasti ja alustatakse otsingut nupuga *Start*.



Seejuures tuleb arvestada järgmiseid seadeid:

Cluster Size: Võimalab määrata sõnade arvu fraasis (selles näites: 2 sõna).

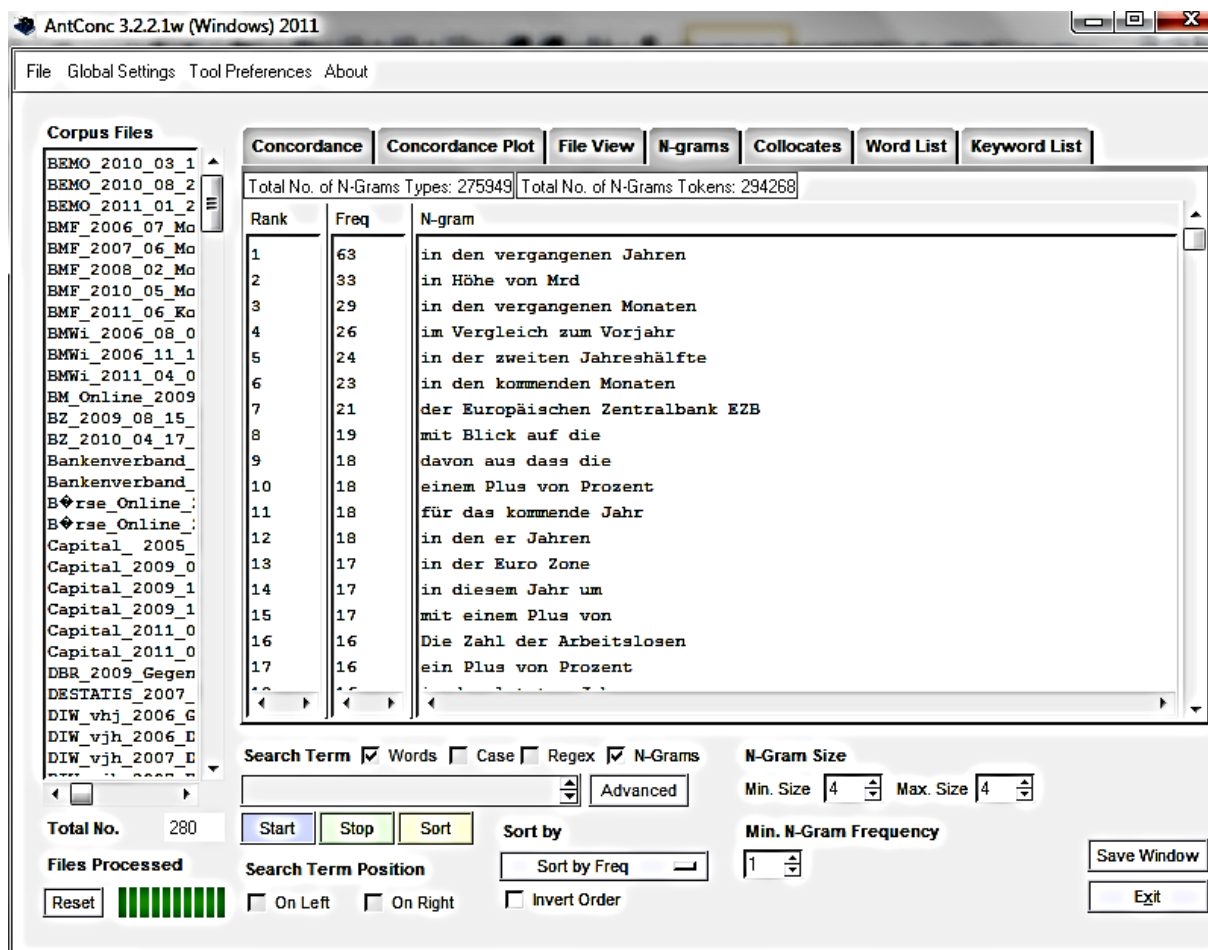
Min. Cluster Frequency: Võimalab määrata, alates millisest esinemise sagedusest fraasi kuvatakse (nt. alates ühest esinemisest).²

Search Term Position: Võimalab määrata, kas otsingusõna asub alati fraasi vasakul või paremal küljel.

b) Päring otsingusõnata

² Üldjuhul võivad sagedamini esinevad tulemused pakkuda suuremat huvi. Kuid see võib sõltuda vastavalt päringu eesmärgist.

Kaartis *Cluster* saab alustada päringut ilma otsingusõnata, märkides ära kasti *N-Grams*. Tulemuseks on tekstikogus sagedasti esinevad sõna rühmad, mille suurust on võimalik kinnitada.



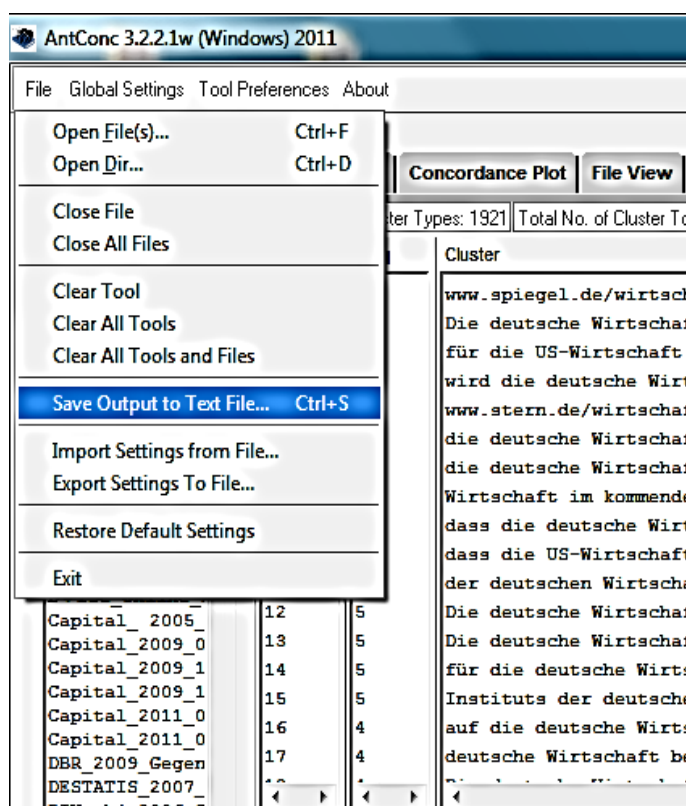
Min./Max. N-Gram Size reguleerib N-grammide suurust. *Min. N-Grams Frequency* määrab N-grammide minimaalset sagedust.³ N-gramme saab sorteerida kas sageduse (*Sort by Freq*) või sõna alguse või lõpu (*Sort by Word/Word End*) järgi.

³ Siinkohal on samuti soovitatav arvestada ainult kõrgeima esinemissagedusega tulemusi.

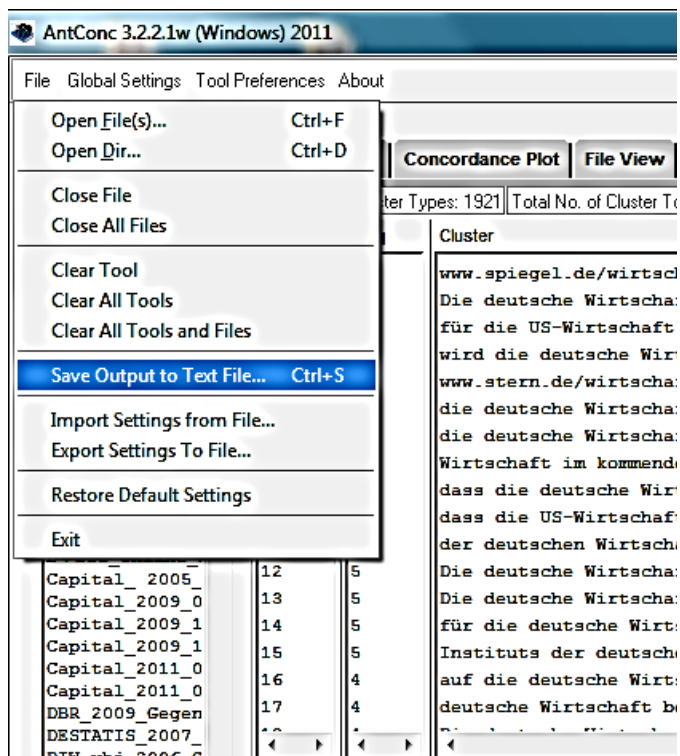
5. TULEMUSTE EKSPORT JA SALVESTAMINE

Pärast analüüsi saab päringute tulemused eksportida ja salvestada arvutisse ning seejärel toimetada neid edasi või kasutada glossaariumi aluseks.

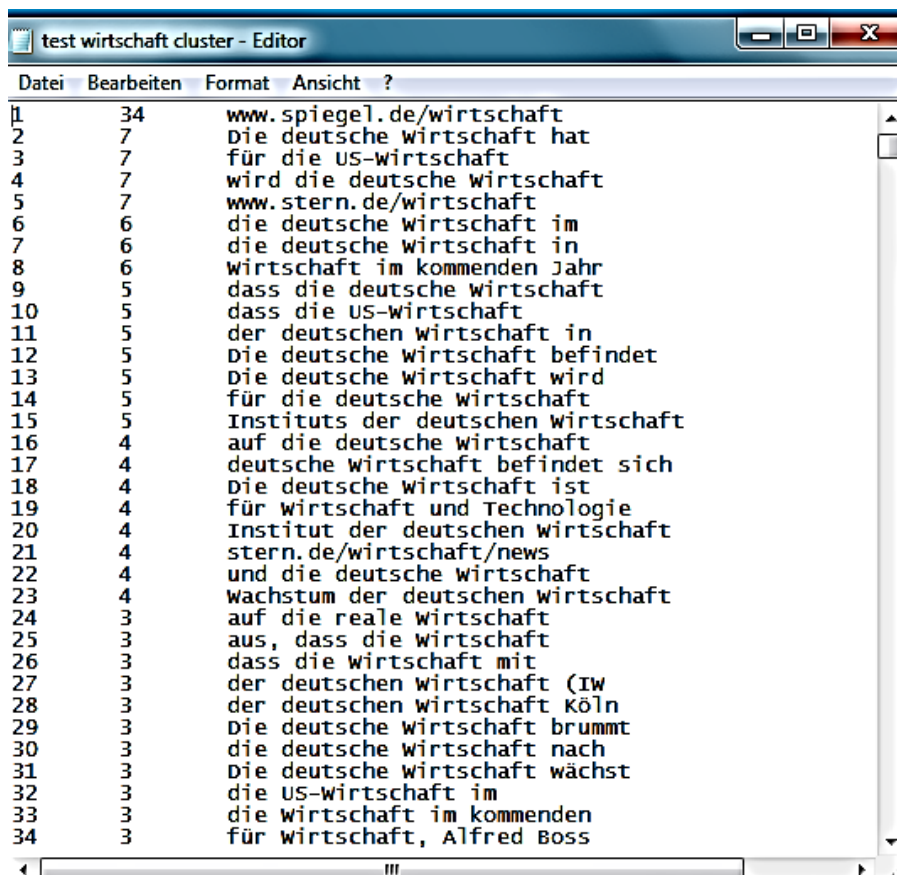
Menüüs **File** saab analüüsi tulemusi salvestada valikuga **Save Output to Text File**. Analüüsi tulemused väljastatakse txt-failina ja neid saab salvestada arvutisse või muule andmekandjale.



Tekstifail sisaldab sõltuvalt päringust sama informatsiooni, mis *AntConci* otsingu aken.



Selles näites on tulemuseks sõnarühmad otsingusõnale *Wirtschaft*. Näidatud on üksiku *Clusteri* järjekord ja sagedus.



Pärast otsinguid on järgmiseks sammuks tulemuste toimetamine tegeliku glossariumi sissekande jaoks, mis võivad vastavalt vajadusele sisaldada erinevaid andmeid. Analüüsiprogramm pakub arvukalt võimalusi ning sõltuvalt teadmiste tasemest ja eesmärgist võivad töötlemise tulemused olla erinevad. Üldiselt on KoGlossi meetodi eesmärgiks täita IV peatükis esitatud kategooriad.

IV. KONSTRUKTSIOONIDE TÖÖTLEMINE GLOSSAARIUMIS

*AntConc*iga väljavalitud konstruktsioonid talletatakse glossaariumi sissekannetena. Glossaariumi sissekandes saab informatsiooni muuta ja kujundada järgmise mustri järgi:

Märksõna või konstruktsioon: ...

Morfoloogia: ...

Teave konstruktsiooni süntaktilise struktuuri kohta: ...

Tähendus(ed): ...

Pragmaatika (kasutus): ...

Korpusest pärinevad näited konstruktsioonist: ...

Muu informatsioon: ...

Seda põhimustrit saavad õpetajad, õpilased või kutselised kasutajad vastavalt oma vajadustele täiendada, lühendada või muuta. Põhimuster sisaldab informatsiooni konstruktsiooni morfoloogia, süntaksi, semantika (täheenduse) ja kasutuse kohta (stiil, teksti tüübi eripära, kasutuse sagedus jne). Sissekanne hõlmab ka korpuse näiteid. Sissekande autoril on võimalus lisada olulisi leide korpuse analüüsist või autori enda teadmistest pärinevat informatsiooni, olenemata sellest, kas on tegu grammatiliste, semantiliste, pragmaatiliste või faktiliselt informatiivsete aspektidega. Sissekande lõppu saab lisada autori nime ja kuupäeva.

Glossaariumi sissekandeid saab luua ja salvestada virtuaalses õpikeskkonnas Moodle, nii nagu on kirjeldatud allpool Moodle peatükis. Valmis glossaariumi sissekanne võib Moodle'is näha välja järgmiselt:

im Vergleich zu/m X:



Morphologie:

Präposition/Artikel_{im} + Substantiv_{sg, Dat.} + Präposition_{zu/m} + X_{Dat.}

X = Nominalphrase, z.B. Artikel + Substantiv, Artikel + Adjektiv + Substantiv

Angabe zur syntaktischen Struktur der Konstruktion:

- **Aufbau:** Präpositionalphrase mit eingebetteter Präpositionalphrase - [im Vergleich [zu/m X]_{pp}]_{pp}
- **Einbettung im Satz:** als Adverbialbestimmung

Bedeutung(en):

- 'in direkter Gegenüberstellung zu einer zweiten Größe'
- **Synonyme:** gegenüber, im Gegensatz zu, verglichen mit, in Relation zu

Gebrauch: wird genutzt als Bindeglied zum Aufzeigen einer Gegenüberstellung oder Verhältnismäßigkeit zweier Größen, vor allem von Zahlenwerten in ihrer zeitlichen Veränderung oder von unterschiedlichen Markt- und Finanzbereichen; kontrastive Funktion;

Verwendungsbeispiele aus dem Korpus:

"Die Zinsausgaben sind abhängig von der Steigerung der Bundesschuld durch die jährliche Nettokreditaufnahme, vom Zinsniveau und von den Konditionen der Anschlussfinanzierungen **im Vergleich zu den Tilgungsleistungen**." (Bundesministerium der Finanzen, Monatsbericht 05/2010)

"Im Schnitt erwarten die Analysten 2010 **im Vergleich zu 2009** einen Gewinnanstieg von 40 Prozent und bis 2011 sogar eine Gewinnverdopplung." (FOCUS Money, 29/12/2009: "Holprige Strecke")

"**Im Vergleich zu deutschen Banken** wachsen die spanischen Großbanken vor allem durch Zukäufe im Ausland." (Handelblatt, 30/12/2005: "Spanische Banken trumpfen auf")

"Nach einem Zuwachs von 5,6 Prozent bei den realen Auftragseingängen in den Betrieben mit 20 und mehr Beschäftigten im Jahr 2007 konnte der Bereich Wirtschaftsbau im Januar **im Vergleich zu dem entsprechenden Vorjahresmonat** ein Plus von gut 20 Prozent realisieren." (Institut der deutschen Wirtschaft Köln, 08/07/2008: "Schwächung ohne Absturz: IW-Konjunkturprognose Frühjahr 2008")

"Zwar ist die Summe der Nettolöhne **im Vergleich zum Vorjahr** im ersten halben Jahr gestiegen, allerdings brachte das dritte Quartal 2010 – **im Vergleich zu den drei Monaten zuvor** – bereits wieder ein Minus von 0,1 Prozent." (ZEIT online, 02/12/2010: "Glaube statt Fakten")

Sonstige Informationen:

Alternativ, aber weniger häufig, wird auch die Konstruktion *im Vergleich mit* X genutzt.

1. GLOSSAARIUMI SISSEKANNETE OSAD

a) Konstruktsioon

Eelnevalt toodud glossaariumi sissekandes on näidatud keeletarkvara *AntConci* abil kindlaks määratud konstruktsioon, mida on kirjeldatud vastavas sissekandes nagu näiteks:

Märksõna või konstruktsioon: *im Vergleich zu X* (võrreldes X-ga)

Teatud keeleliste mustrite ja konstruktsioonide leidmiseks, mis on omased teatud eriala- või võõrkeelele, saab *AntConci* mitmeti kasutada: esiteks saab funktsiooniga *Cluster* (vt III. peatükk) luua nimekirju, mis sisaldavad sagedasti koos esinevaid sõnu. Otsinguga *Collocates* on võimalik leida ühe kindla väljendiga seotud sõnu (vt. III. peatükk). Mõlemal juhul tuleb esimesi otsingutulemusi pidada esialgseteks. Sihipärase uurimisega tuleb kontrollida, kas leitud sõnarühmadel on tõepoolest tähendus ja kas leitud sõna muster on sellele valdkonnale tõesti omane, näiteks majanduskeel, teaduskeel, reklaami disainerite keel, jne. Seejuures on otstarbekas uurida lähemalt leitud sõnade ja sõnade

rühmade kontekste KWIC-ridade (funktsiooni *Concordance*) ja funktsiooni *FileView* abil.

Tuleb märkida, et on juhtumeid, kus osade sõnade rühmade sees võivad esineda muutujad, ehk teatud elemendid, mis ei ole fikseeritud ühele sõnale. Sellised mustreid saab lisada glossaariumisse kasutades kohatäidet ‚X‘ (vt näide eespool).

b) Teave morfoloogiast

Teavet morfoloogia (informatsioon sõnaliigi ja muutelõpu kohta) kohta on mõistlik lisada, siis kui meetodit kasutatakse keeleteaduslikel õpingutel. Seejuures võib loetleda konstruktsiooni koostisosade **sõnaliigid**, näiteks *Bruttosozialprodukt* (nimisõna) ja teavet **pöördeparadigma** kohta (Pl.: *die Bruttosozialprodukte*, Gen.: *des Bruttosozialprodukt(e)s*, jne). Samuti võivad kuuluda siia üksikasjad **sõnamoodustusest**. Selleks et näidissõna ehitust selgelt edasi anda võib seda nii eraldada: *Brutto + sozial + produkt*.

Teavet konstruktsiooni morfoloogia kohta saab laiendada, nt. *im Vergleich zu X*

Morphologie:

Präposition/Artikel_{im} + Substantiv_{Sg., Dat.} + Präposition_{zu/m} + X_{Dat.}

X = Nominalphrase, z.B. Artikel + Substantiv, Artikel + Adjektiv + Substantiv

Siinkohal saab kasutaja teada, milline vorm on konstruktsiooni komponentidel (esimene rida), mis võib muutuja elemendi (=X) kohale tulla (teine rida). Sõnade ja sõnarühmade erandvorme on võimalik leida kohatäitjate (*Wildcards*) abil (vt III peatükk).

c) Teave konstruktsiooni süntaktilisest ehitusest

Selles punktis seletatakse millistest elementidest konstruktsioon koosneb ja millist funktsiooni see analüüsitavas tekstis tavaliselt täidab. Seejuures on võimalik vabalt valida, milline vormistus mingile sihtgrupile kõige paremini sobib.

Konstruksiooni *im Vergleich zu X* võib näiteks liigitada ehituse järgi prepositsiooniliseks fraasiks ja lause funktsiooni järgi määruseks.

Angabe zur syntaktischen Struktur der Konstruktion:

- **Aufbau:** Präpositionalphrase mit eingebetteter Präpositionalphrase - [im Vergleich [zu/m X]_{PP}]_{PP}
- **Einbettung im Satz:** als Adverbialbestimmung

Vajalikku informatsiooni saab kätte, vaadates sõna mustrite juurde kuuluvaid KWIC-ridasid (funktsioon *Concordance*) ja leides ning seejärel kategoriseerides erinevad kasutamise kontekstid ja funktsioonid.

Töölases tegevuses võib sissekandel olla võimalikult lihtne vorm, st see sisaldab informatsiooni grammatika kohta ainult minimaalselt, nagu näiteks:

im Vergleich zu X:

Grammatik: Vergleich, der; im Vergleich zu + Dativ

d) Teave tähendusest

Teave tähenduse (semantika) kohta on oluline nii keeleteaduse õppe valdkonnas kui ka professionaalsetele kasutajatele. Tähendus(te) all mõistetakse konstruktsiooni ümberkirjutamist (parafraseering). Tähendus tuleneb korpusest. Parafras „otsene vastandamine mingi teise suurusega“ kirjeldab näiteks konstruktsiooni *im Vergleich zu* demantilist sisu. Samuti on kirjeldamisel abiks samasuguse või sarnase tähendusega sõnad (sünonüümid) (toodud näites *gegenüber*, *im Gegensatz zu*, jne), vastandtäenduslikud sõnad (antonüümid). Konstruktsiooni *im Vergleich zu X* tähendust võib kirjeldada järgmiselt:

Bedeutung(en):

- 'in direkter Gegenüberstellung zu einer zweiten Größe'
- **Synonyme:** *gegenüber*, *im Gegensatz zu*, *verglichen mit*, *in Relation zu*

Sõnade ja sõnarühmade erinevaid tähendusi on võimalik tekstikogust pärit otsingutulemuste abil lahti seletada ja üldistada. Siinjuures on samuti võimalik kasutada KWIC-ridasid ja täisteksti vaadet (*FileView*).

e) Teave kasutuse kohta

Teave kasutusest (pragmaatika) sisaldab informatsiooni stiili, konstruktsiooni kasutuse sageduse jms kohta. Seejuures võivad olla oluline teksti tüüp või isegi konkreetse teema või domeeniga piirdumine.

Konstruktsiooni *im Vergleich zu X* kasutamise kirjelduses selgitatakse näiteks, millise funktsiooniga see konstruktsioon tekstides esineb ja milliste kontekstidege see on peamiselt seotud:

Gebrauch: wird genutzt als Bindeglied zum Aufzeigen einer Gegenüberstellung oder Verhältnismäßigkeit zweier Größen, vor allem von Zahlenwerten in ihrer zeitlichen Veränderung oder von unterschiedlichen Markt- und Finanzbereichen; kontrastive Funktion;

Lõpuks võib märkida, millist konkreetset hoiakut sõna/konstruktsioon kõnelejas väljendab (negatiivset või positiivset hinnangut, tahtlikkut kaugenemist objektist jt). Konstruktsioon *Negativwachstum* ehk negatiivne kasv kasutatakse selleks, et negatiivset arengut keeleliselt ilustada.

Teave konstruktsiooni kasutamise kohta on oluline nii keeleteaduse õppijatele kui ka professionaalsetele kasutajatele.

f) Korpusest pärinevad näited konstruktsiooni kasutuse kohta

Tekstikogust pärinevad näiteid konstruktsiooni kasutusest näitavad ühelt poolt, millises kontekstis teatud konstruktsioon esineb, ehk milline on selle süntaktiline vorm ja funktsioon. Teisalt on autentsete näidetega parem illustreerida konstruktsiooni tähendust ja kasutust. Näited kasutuse kohta võetakse *AntConc*i abiga korpusest ja lisatakse sissekandesse. Selleks valitakse umbes viis täislauset, millel oleks võimalikult erinevad

kontekstid ja funktsioonid ning vahetevahel esinevad tähenduse variatsioonid. Konkreetse näite tuvastamise huvides võib lisada sulgudesse selle allika.

Verwendungsbeispiele aus dem Korpus:

"Die Zinsausgaben sind abhängig von der Steigerung der Bundesschuld durch die jährliche Nettokreditaufnahme, vom Zinsniveau und von den Konditionen der Anschlussfinanzierungen **im Vergleich zu den Tilgungsleistungen.**" (Bundesministerium der Finanzen, Monatsbericht 05/2010)

"Im Schnitt erwarten die Analysten 2010 **im Vergleich zu 2009** einen Gewinnanstieg von 40 Prozent und bis 2011 sogar eine Gewinnverdopplung." (FOCUS Money, 29/12/2009: "Holprige Strecke")

"**Im Vergleich zu deutschen Banken** wachsen die spanischen Großbanken vor allem durch Zukäufe im Ausland." (Handelblatt, 30/12/2005: "Spanische Banken trumpfen auf")

"Nach einem Zuwachs von 5,6 Prozent bei den realen Auftragseingängen in den Betrieben mit 20 und mehr Beschäftigten im Jahr 2007 konnte der Bereich Wirtschaftsbau im Januar **im Vergleich zu dem entsprechenden Vorjahresmonat** ein Plus von gut 20 Prozent realisieren." (Institut der deutschen Wirtschaft Köln, 08/07/2008: "Schwächung ohne Absturz: IW-Konjunkturprognose Frühjahr 2008")

"Zwar ist die Summe der Nettolöhne **im Vergleich zum Vorjahr** im ersten halben Jahr gestiegen, allerdings brachte das dritte Quartal 2010 – **im Vergleich zu den drei Monaten zuvor** – bereits wieder ein Minus von 0,1 Prozent." (ZEIT online, 02/12/2010: "Glaube statt Fakten")

Teave konstruktsiooni kasutusest on oluline nii keeleteaduses kui ka tööalaselt.

g) Muu informatsioon

Selles kirjelduses on sissekande autoril võimalik kirjutada märkusi konstruktsiooni erinevate aspektide kohta, mis ei sobi muu teave juurde. Näiteks võib konstruktsiooni (*ein/der*) *Zuwachs von X Prozent* juurde lisada järmist:

Sonstige Informationen: Alternativ zum Wort "Prozent" wird in der Konstruktion häufig das Zeichen "%" benutzt.

Järgnevas peatükis on kirjeldatud glossaariumi sissekannete tometamine virtuaalses õpikeskkonnas Moodle.

V. MOODLE

1. MIS ON MOODLE?

Moodle on nn. LMS (Learning Management System). See võimaldab kursuste ja online-koolituste haldamist ja läbiviimist. Seejuures on võimalik täiendada õppetegevuse sisu. Kogu suhtlus võib näiteks toimuda süsteemi kaudu ja võimalik on jagada täiendavaid materjale, ülesandeid, teste.

Kuna Moodle on võrgurakendus, siis peab see olema vastavas asutuses (kool, ülikool, ettevõtte) tsentraalselt installeeritud ja internetis kättesaadav. Rohkem informatsiooni Moodle'i kohta saab järgmisel aadressil: <http://moodle.org>.

Kui Moodle on juba kättesaadav, tuleb kohalikul Moodle'i administraatoril lasta luua tühi kursuse põhi, kusjuures vajalikud on koolitaja või õpetaja õigused (vt. peatükki Glossaariumi haldamise kohta „Glossaariumi rollimäärang“).

Moodle'it on võimalik mitmeti kasutada. Vastloodud kursuses on võimalik korraldada erinevaid õppetegevusi, sealhulgas õppetegevust „Glossar“.

„Glossar“ kui õppetegevus Moodle's võimaldab luua, toimetada ja laiendada sõnade katalooge. Registreeritud õpetaja saab oma kursuses luua ja hallata õppetegevust „Glossar“.

Järgnevalt tutvustatakse põhiaidmeid glossaariumi õppetegevuse koostamise ja kasutamise kohta.

2. GLOSSAARIUMI LOOMINE

Sisenedes esimest korda Moodle'i kursuse ruumi on see „tühi“ (joon. 1). Kõik kursuse tegevused ja materjalid tuleb koostada.

Moodle'i kursuse ruum koosneb **plokkidest** (kitsad veerud vasakul) ja **kursuse lõikudest** (keskmised laiad veerud).

Plokid „*Navigation*“ (navigatsioon) ja „*Einstellungen*“ (seaded) võimaldavad kursuses navigeerida.

Kursuse lõikudes hakkavad edaspidi olema konkreetsed õppesisud. Sinna saab lisada „õppematerjale“ ehk *Arbeitsmaterialien* ja „tegevusi“.

Õppematerjalide all mõeldakse erinevaid elektroonilisi õppesisusid nagu näiteks mistahes formaadis failid, lingid, andmebaasid jms. Õppetegevused on tegevused, mida õppijatel tuleb kursuse käigus sooritada, nagu näiteks glossaariumi koostamine.

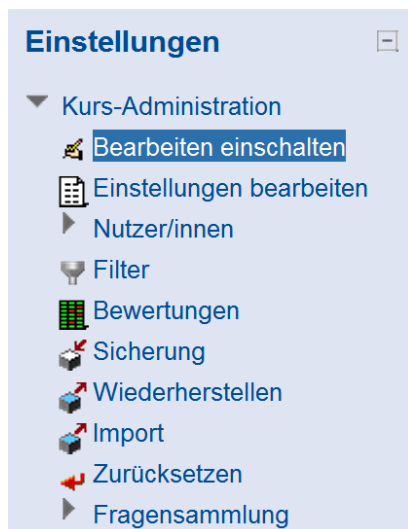
The screenshot displays two side-by-side panels from a Moodle course interface. The left panel contains two main sections: 'Navigation' and 'Einstellungen'. The 'Navigation' section lists 'Startseite' with sub-items 'Meine Startseite', 'Website', 'Mein Profil', and 'Meine Kurse'. The 'Einstellungen' section is expanded, showing 'Kurs-Administration' with sub-items: 'Bearbeiten einschalten', 'Einstellungen bearbeiten', 'Nutzer/innen', 'Filter', 'Bewertungen', 'Sicherung', 'Wiederherstellen', 'Import', 'Zurücksetzen', 'Fragensammlung', and 'Kursdateien'. Below these are 'Rolle wechseln...' and 'Mein Profil'. The right panel, titled 'Themen dieses Kurses', shows a list of eight topics, each with a number and a checkbox on the right.

Themen dieses Kurses	
1	<input type="checkbox"/>
2	<input type="checkbox"/>
3	<input type="checkbox"/>
4	<input type="checkbox"/>
5	<input type="checkbox"/>
6	<input type="checkbox"/>
7	<input type="checkbox"/>
8	<input type="checkbox"/>

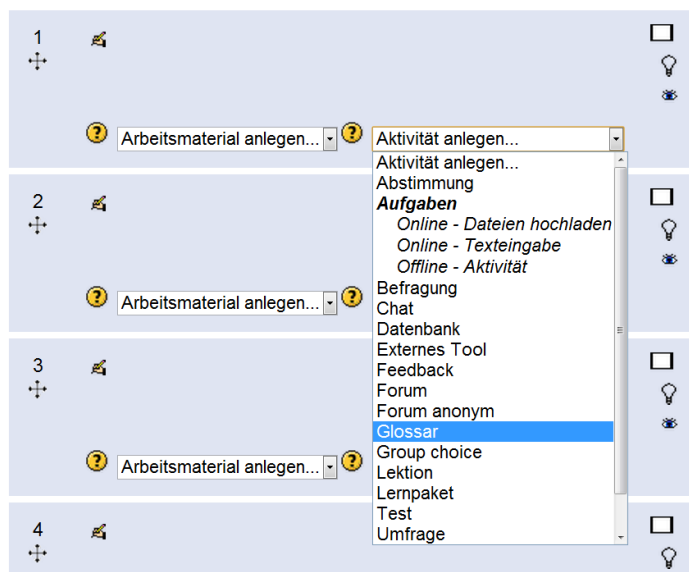
Joonis 1: Kursuse ruum

Moodle's glossaariumi loomiseks tuleb teha järgmisi samme:

a) Ploki „*Einstellungen*“ (seaded) all asub rubriik „*Kurs-Administration*“ (kursuse administreerimine). Selle all on võimalik lülitada sisse režiim „*Bearbeiten einschalten*“ (luba muutmise) (joonis 2).



Joonis 2



Joonis 3: Glossariumi loomine

b) Ilmuvad ikoonid (nt küsimärk jne), mis suunavad otse nendesse kursuse kohtadesse,

c) mida on võimalik muuta.

d) Paremal kursuse lõikudes ilmuvad kaks menüüd: „*Arbeitsmaterial anlegen*“ (lisa õppematerjal) ja „*Aktivität anlegen*“ (lisa tegevus). Valige „*Aktivität anlegen*“ all olev punkt „*Glossar*“ (joonis 3).

e) Järgmise sammuna saab vastavalt oma eelistustele teha seadistusi glossaariumi jaoks, näiteks „*Kommentare*“, mis võimaldab kirjutada kommentaare sissekannete juurde või „*Automatisch verlinken*“, mis seob sissekanded omavahel. Kollased esiletõstetud küsimärgid annavad täpsemat lisateavet iga seade kasutamise kohta. Glossaariumile tuleb anda nimetus ja kirjeldada selle eesmärki. Pärast seadete lõplikku salvestamist, suunatakse tagasi kursuse juurde.

2.1. GLOSSAARIUMI SISSEKANDE LOOMINE

Pärast glossaariumi loomist on nüüd võimalik glossaariumi pealeheküljel koostada sissekandeid. Klõpsates nupule „*Eintrag hinzufügen*“ (lisa sissekanne) avaneb uus aken, kuhu saab kirjutada sissekande (joonis 4):

The screenshot shows a web form titled "Glossar" with a yellow header. The form is divided into several sections:

- Grundeinträge**: Contains a text input field labeled "Begriff*" (Concept).
- Definition***: Includes a rich text editor with a toolbar for font style (bold, italic, underline), font size, paragraph alignment, bulleted and numbered lists, indentation, link, unlink, and image insertion. Below the editor is a text field labeled "Pfad: p" (Path).
- Alternativbegriffe**: A text input field with a help icon.
- Anhang**: A section for attachments with a "Datei hinzufügen" (Add file) button, a note "Maximale Dateigröße: 400MB", and a status bar showing "Keine Datei" (No file).

Joonis 4

- Soovitud väljend (termin või konstruktsioon) tuleb kirjutada kasti „*Begriff*“ (mõiste).
- Tekstikastis saab punkti haaval täita artikli struktuuri. Glossaariumi sissekandesse lisatav teave peaks olema koostatud läbimõeldud mustri järgi (artikli struktuuri

üksikasjade kohta vt peatükk IV). Seda on hiljem võimalik kohaldada mallina glossaariumis. Funktsioonis toimetaja on saadaval mitmeid tekstivormindusfunktsioone.

c) Igale glossaariumi sissekandele saab lisada „*Alternativbegriffe*“ (sünonüüme, võtmesõnu). Mahukamate glossaariumide puhul hõlbustab see nende otsingut.

d) Selleks et artiklid oleksid omavahel seotud, tuleb aktiveerida automaatne sidumine, klõpsates nupule „*Einträge automatisch verlinken*“ (joonis 5).

e) Iga glossaariumi sissekande puhul on võimalik aktiveerida individuaalselt *Auto-Verlinkung* all valikud „*Groß-/Kleinschreibung*“ ja „*Nur vollständige Worte*“.

f) ***Groß-/Kleinschreibung***. See seade määrab, kas automaatse sidumise puhul arvestatakse suur- ja väiketähtedega.

g) ***Nur vollständige Worte***. Kui see valik on sisse lülitatud, seotakse sissekanne ainult siis, kui tekstis olev täielik sõna ühtib glossaariumi sissekandega. Seost ei tekitata siis, kui näiteks glossaariumis on tehtud sissekanne „*Konstrukt*“, kuid tekstis on hoopis sõna „*Konstruktivität*“.

h) Seejärel tuleb valminud artikkel salvestada.

2.2. LOODUD ARTIKLITE TOIMETAMINE

a) Kui glossaariumi sätetes on valik „*Immer bearbeitbar*“ märgitud „*Ja*“ peale, saavad osalised alati muuta oma sissekandeid.

b) Sissekande muutmiseks tuleb klõpsata käe sümbol peale.

c) Sissekande kustutamiseks tuleb klõpsata risti ikooni peale.

2.3. KOMMENTAARIDE KOOSTAMINE

Funktsiooni „*Kommentare*“ abil saavad osalejad ja õpetajad avaldada oma arvamust, teha ettepanekuid või jagada oma kogemusi valmis sissekannete kohta, nuu nagu all

toodud näites saksa KoGlossi glossaariumist (milles dotsent proovib motiveerida ja soovib parandust):



Kommentaariid edendavad koostööd osalejate vahel ning seeläbi aitavad kaasa glossaariumi kvaliteedi parandamisele, eriti siis, kui osalejad kommenteerides omavahel suhtlevad.

Kommentaari koostamiseks:

- a) Vajutage lingile „*Kommentare*“,
- b) sisestage tekst ediitori tekstiväljale ja seejärel salvestage.

Olemasolevad kommentaarid on nähtaval nimekirjana glossaariumi sissekande all koos autori ja kuupäevaga.

3. OLEMASOLEVATE GLOSSAARIUMITE KASUTAMINE

3.1. GLOSSAARIUMI VAATAMINE

Olemasolevat glossaariumit saab vaadata ja sorteerida tähestikuliselt, kuupäeva või autori järgi (joonis 6):

Joonis 6

3.2. GLOSSAARIUMI OTSIMINE

Glossaariumi otsingu alustamiseks kasutage kasti „Suchen“(otsi).

Kui soovite teostada otsingut ainult mõistete hulgas, tuleb seade „Volltext-Suche“ otsingukastist paremal välja lülitada.



Kui valik „Volltext-Suche“ on aktiveeritud, siis otsitakse sõna mitte ainult mõistete hulgast, vaid terve glossaariumi tekstist.

3.3. GLOSSAARIUMI IMPORT/EKSPORT

Moodle's on võimalus kasutada juba loodud glossaariume. Teistest kursustest pärit glossaariume on võimalik importida oma kursusesse või salvestada kursuse glossaarium ekspordi funktsiooni kaudu.

Glossaariumi eksport:

a) Esmalt tuleb avada glossaarium, mida soovitakse eksportida, seejärel klõpsata „Einstellungen“ all oleva „Glossar-Administration“ ja lõpuks „Glossar exportieren“ peale (Joonis 7).



Joonis 7

b) Klõpsates „*Glossar in Datei exportieren*“ (joonis 8) peale, avab see pop-up akna, kus tuleb valida „*Datei speichern*“ (salvestamise asukoht) ja „Ok“. Glossaarium salvestatakse lokaalselt XML-failina. Seda faili on igal ajal võimalik laadida jälle üles Moodle'i impordi funktsiooni kaudu.

Glossar als XML-Datei exportieren

Glossar in Datei exportieren

Joonis 8

Glossaariumi import:

a) Ühte temaatilisse plokki omal valikul, tuleb luua uus glossaarium, valides menüüs „*Arbeitsmaterial anlegen*“ all „*Glossar*“.

b) Esmalt tuleb avada vastloodud glossaarium ja seejärel valida „*Einstellungen*“ all *Glossar-Administration* ja klõpsata siis „*Einträge importieren*“ (impordi sissekanne).

Einträge aus XML-Datei importieren

Datei importieren ?

Datei wählen ...

Keine Datei

Ziel für importierte Einträge ?

Aktuelles Glossar ▾

Kategorien importieren

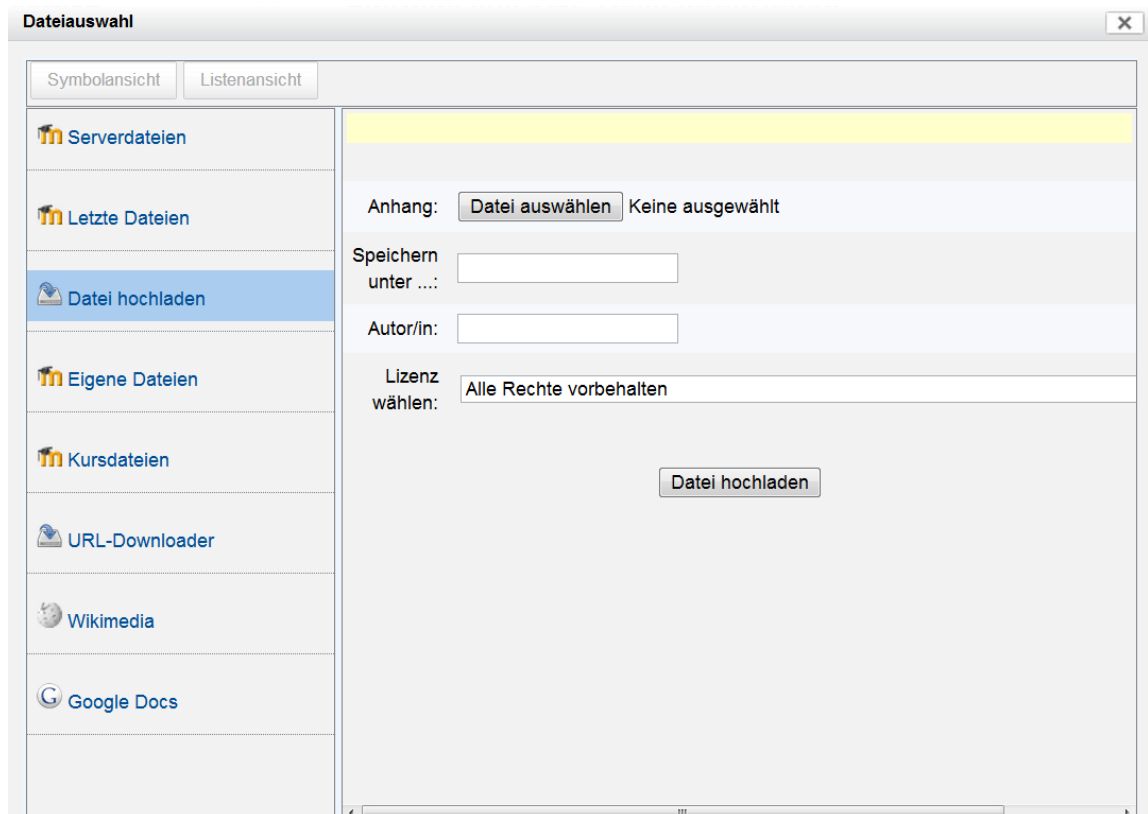


Speichern

Joonis 9

c) Vajutades „*Datei wählen*“ (vali fail) (Joonis 9) peale avab see pop-up akna (joonis 10).

d) Pop-up aknas tuleb valige „*Datei hochladen*“ ja klõpsata „*Datei auswählen*“ peale ning leida arvutist alguses salvestatud fail (vt Glossaariumi eksport, punkt II).



Joonis 10

e) Seejärel valige „Datei hochladen“.

f) Seejärel salvestage üleslaaditud fail.

3.4. GLOSSAARIUMI HALDAMINE. ROLLIMÄÄRANG

See kuidas kasutaja Moodle's tegutseda ja hallata saab, sõltub nn. kasutaja rollimäärangust. Rollid määratleb Moodle administraator. Neid on võimalik valikuliselt muuta. Peamised määratavad rollid ühes Moodle kursuses on järgmised:

Õpetajad

Õpetajad haldavad kursust ja saavad seda sisustada vastavalt oma vajadustele. Õpetajatel on võimalik võtta teisi rolle (nt osaleja, et näha kursuse sisu nende vaatenurgast). Õpetajatel on õigus määrata teistele osalejate rolle. Kuid nad peavad olema kursusele registreeritud.

Rolli saab määrata järgmiselt:

a) Glosaariumi seadetes vajutada „*Lokale Rollen zuweisen*“ ja valida määratav roll, nt õpetajad (joonis 11).

Wählen Sie bitte eine Rolle zur Zuweisung		
Rolle	Beschreibung	Nutzer/innen mit einer Rolle
Lehrende(r)	Trainer/innen dürfen in einem Kurs alles tun, auch Aktivitäten bearbeiten und Teilnehmer/innen beurteilen	0
versteckte(r) Lehrende(r)	Die Rechte stimmen mit denen der gewöhnlichen Lehrenden überein; allerdings sollen Nutzer mit dieser Rolle nicht in der Kursübersicht als Lehrende angezeigt werden.	0
Kontrollierende	Kontrollierende haben die gleichen Berechtigungen wie Trainer/innen ohne Bearbeitungsrecht, können aber in der Bewerberübersicht die Bearbeitung einschalten.	0
Lehrende(r) ohne Bearbeitungsrecht	Trainer/innen ohne Bearbeitungsrecht dürfen in Kursen unterrichten und Teilnehmer/innen bewerten, aber sie können nichts verändern	0

Joonis 11

b) Seejärel ilmuvad kaks nimekirja: „*Vorhandene Nutzer/innen*“, kellel on juba õpetaja roll ja „*Potenzielle Nutzer/innen*“, kellele Teil on võimalik määrata õpetajate rolli.

c) Valige paremast nimekirjast välja potentsiaalse(d) kasutaja(d) ja vajutage seejärel nupule „*Hinzufügen*.“ Sellega olete määranud valitud isiku(te)le õpetaja rolli.

Osalejad (õpilased)

Osalejatel on võimalik kasutada kursust ja selles sisalduvaid tegevusi, õppematerjale ja plokke, kuid nad ei saa muuta kursuse seadeid.

Külaline

Külalistel on võimalik vaadata kursuse sisusid, kuid nad ei saa neid kasutada või muuta. Selleks tuleb esmalt kursuse seadete all aktiveerida juurdepääs külalistele.

VI. RAKENDAMINE HARIDUSES JA TÖÖALASES TEGEVUSES

KoGlossi meetod koosneb järgmistest sammudest:



Korpuse
koostamine

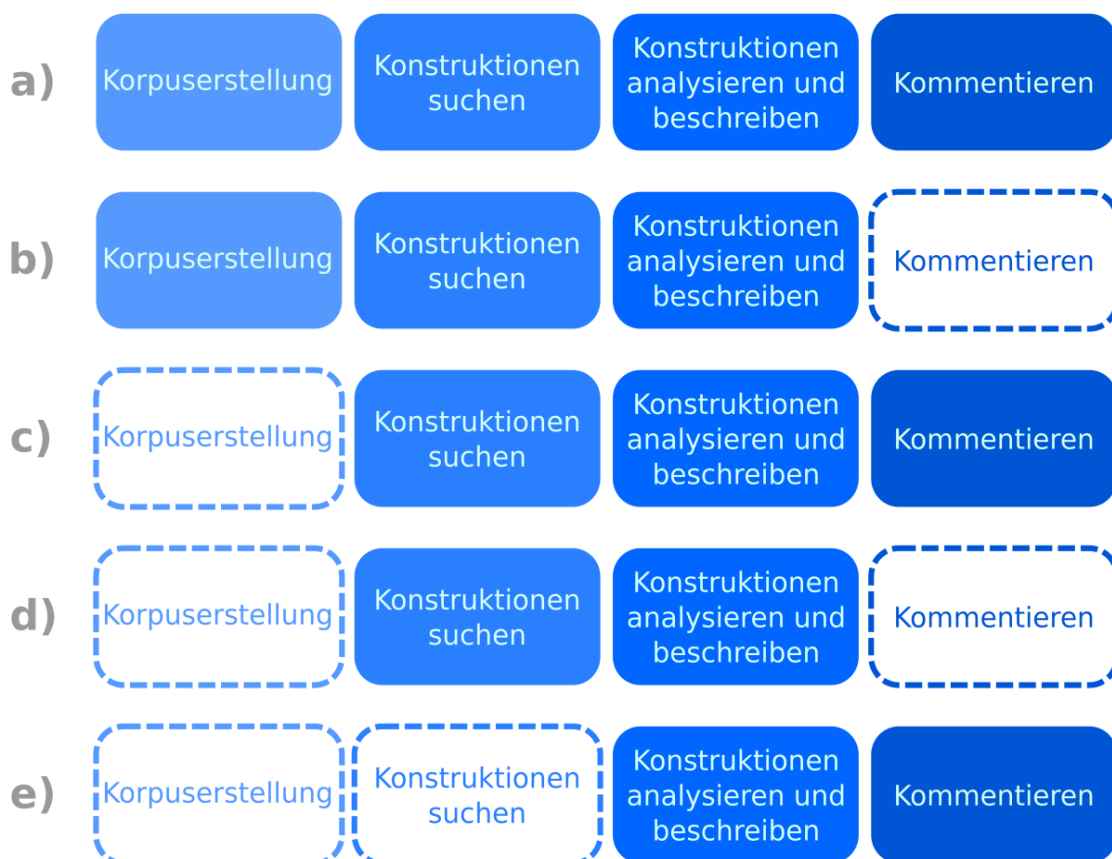
Konstruksioonide
otsimine

Konstruksioonide
analüüsimine ja
kirjeldamine

Kommenteerimine

Vastavalt sihtgrupile on võimalik meetodit kasutada nii tervikuna (a) kui ka üksikute sammudena (b, c, d, e).

KoGloss meetodi rakendusvõimalused



Iga sihtrühma puhul on võimalik otsustada töösammude kombinatsiooni üle, mis vastab nende vajadustele ja soodustab saavutatud oskusi kõige paremini.

Rakendusvõimalus	Potentsiaalsed sihtgrupid	Saavutatavad pädevused
a	Heade keeleteaduslike eelteadmisetega üliõpilased, (nt tõlkeõpetus ja muud filoloogilised erialad), keeleõpetajad	Võõrkeele ja/või emakeele pädevus, erialakeele pädevus, üldteaduslik(metoodiline) pädevus, teksti lingvistiline pädevus, leksikograafiline pädevus, keele õpetamise pädevus, lingvistiline kommunikatsioon
b	Keeleteaduslike eelteadmisetega üliõpilased, keeleõpetajad, professionaalsed kasutajad (tõlkijad, tõlgid, terminoloogid, ajakirjanikud jt)	Võõrkeele ja/või emakeele pädevus, erialakeele pädevus, teksti lingvistiline pädevus, leksikograafiline pädevus
c	Keeleteaduslike eelteadmisetega üliõpilased, keeleõppijad, kasutajad (ajakirjanikud, suhtekorraldajad, sekretärid jne)	Võõrkeele ja/või emakeele pädevus, erialakeele pädevus, lingvistiline kommunikatsioon
d	Keeleõppijad, professionaalsed kasutajad	Võõrkeele ja/või emakeele pädevus, erialakeele pädevus,
e	Keeleteaduslike eelteadmisetega üliõpilased, keeleõppijad	Võõrkeele ja/või emakeele pädevus, erialakeele pädevus, lingvistiline kommunikatsioon

Korpuse koostamise põhimõtteid võib selgitada igale sihtgrupile. Korpuse koostamine on eriti soovitatav tõlkeõpetuse üliõpilastele, et harjutada informatsiooni ja teksti otsingut. Lühikursuste või projektide puhul võib kasutada juba eelnevalt koostatud tekstikorpust.

Konstruksioone võivad esitada õpetajad, st õppijad töötavad ainult analüüsi/kirjeldusega ja kommenteerivad valmis sissekandeid (e).

KoGloss meetodit saab hästi kasutada ka individuaalselt, näiteks tõlkijate/tõlkide infootsingus ja konkreetsete ning kestvalt kasutatavate informatsiooni allikate loomisel. Selle tarbeks on võimalik koostada ise korpuse (b) või kasutada mõnda olemasolevat

(d). Kui meetodit kasutatakse rühmades, siis tuleks esialgu välja töötada ühtne glossaariumite sissekannete vormistus, et osalejad saaksid neid iseseisvalt koostada. Vajadusel tuleks õpetada/korrata keelelist terminoloogiat.

Lõpp-produkt, st loodud glossaariumid leiaksid kasutust käsitletud keele õpetamisel ja omandamisel, nii esimese kui ka teise võõrkeelena, teksti koostamises emakeeles ja võõrkeeles, tõlkimise harjutustes, terminoloogia õpetamises, kursustel, mis edendavad teaduslikku pädevust, jne. Üks- ja mitmekeelseid konstruktsioone saab uurida kodu-, semestri- ja bakalaureusetöös, mis on pühendatud lingvistilisele ja tõlketeaduslikule temaatikale.

Konstruktsioonide glossaariume saab kasutada sõltuvalt keelelise pädevuse vajadusest kutsetegevuses. Esmajoones saab neid kasutada ka ainealases teksti tootmises. Sõltuvalt valdkonnast saab teha kindlaks erialase diskursuses levinud konstruktsioonid ja kasutada neid erialakeelega lõimitud tekstide koostamises. Eriti kasulikud oleksid sellised erialakeelsed konstruktsioonide glossaariumid algajatele, kelle igapäevases töös on oluline kirjalik suhtlemine (sekretärid, asjaajajad jm). Siinjuures on olulised glossaariumid nii emakeeles kui ka võõrkeeles (näiteks suheldes välismaiste äripartneritega).

Konstruktsioonide glossaariumide kasutamise sihtrühma kutsetegevuses kuuluvad ka tõlkijad ja tõlgid. Mingis erialas kasutatavat sõnavara on võimalik koondada tervete konstruktsioonide kujul, ning seega vastavalt nende diskursuse sihipärasele kasutusele, ühte glossaariumisse. Sellise lähenemise abil saaksid peamiselt tõlgid mitte ainult üksikuid erialalisi termineid, vaid ka vastavas valdkonnas kasutatavaid fraase omandada ja nende osas kohaneda. Ka sel juhul võivad konstruktsioonide glossaariumid olla kasulikud nii ema- kui ka võõrkeeles.

2. Magistritöö teoreetilis-metodoloogilised lähtekohad

Oma magistritöös tuginesin funktsionaalsele tõlketeooriale, milles käsitletakse tõlget sihtkultuuri lahutamatu osana ja oluliseks peetakse eelkõige tõlke kommunikatiivset eesmärki. Tõlge ei pea tingimata järgima täpselt lähteteksti, vaid täitma selle funktsiooni. Funktsionaalse tõlketeooria eeliseks on see, et pakutakse teoreetilist raamistikku, mis aitab lahendada tarbetekstide tõlkimisel tekkivaid probleeme ja arvestatakse sihtkultuuri iseärasustega. (vrd Loik 2005) Üks tuntumaid funktsionaalseid teooriaid on skopos-teooria, millele pani aluse Hans Josef Vermeer (24. september 1930 Iserlohn kuni 4. veebruar 2010 Heidelberg), kes oli üks juhtivaid saksa teoreetikuid. Ta avaldas selle koos Katharina Reißiga 1984. aastal raamatus "*Grundlegung einer allgemeine Translations theorie*" (vrd Kask 2011: 6) Seejärel arendas Christiane Nord Skopos-teooria põhireegleid hiljem edasi. Nordi eesmärgiks oli luua lähteteksti analüüsimudel, mis ei sõltuks tekstiüübist, lähte- või sihtkeelest, tõlkija pädevusest ega tõlkesuunast. Nordi analüüsimudel jaguneb tekstisisestest ja -välisest tegurite ning mõju analüüsiks. Seejuures on keelevälised ja sisulised tegurid omavahel seotud ja mõjutavad üksteist. (vrd Saluäär 2007) Nordi analüüsimudelit saab võtta kokku järgmiste küsimustega:

„kes edastab, mis teemal ta ütleb, kellele, mida, milleks, (mida mitte?), missugusel teabekandjal, mis järjekorras, kus, missuguste mittekeeleliste elementidega, millal, mis sõnadega, miks, missuguste lausetega, mis funktsiooniga teksti, mis toonil, missuguse mõjuga?“ (Saluäär 2007)

Lingvistilise tõlkekäsitluse (sõnasõnalise tõlke) üks esindajaid oli keele- ja tõlketeadlane Wolfram Wilss (25. juuli 1925 Ravensburg kuni 3. august 2012 Saarbrücken). Wilssi kohaselt on edukas tõlkimine enam-vähem süstematiseeritud probleemide lahendamise strateegiate ja otsustusprotsesside tulemus. Tema arvates ei ole tõlketeaduses alati võimalik olla ratsionaalne, tuleb arvestada ka loominguilisuse ja intuitsiooniga. (Wilss 1988: IX)

KoGlossi manuaal tõlkimisel oli eesmärgiks, et minu tõlge täidaks sama funktsiooni nagu selle saksakeelne lähtetekst. Tõlge peab olema korrelatsioonis lähtetekstiga. (Reiß, Vermeer 1991: 119) Nii lingvistiline kui ka funktsionaalne suund tegelevad igapäevakeele ja tarbetekstidega. Lähteteksti puhul on tegu tarbetekstidega, mis peaks

tutvustama KoGlossi meetodit. Selles antakse juhiseid, kuidas koostada korpuseid, otsida nendest konstruktsioone ja koostada sissekandeid. Tutvustatakse ka selleks vajaminevaid vahendeid nagu keeletarkvara *AntConc* ja Moodle'i keskkond. Selle tekstiga pakutakse põhimõtteliselt kellegile informatsiooni ehk *Informationsangebot* (Reiß, Vermeer 1991: 119) Kuna selle meetodi potentsiaalsed kasutajad ei pea omama erilisi keeleteaduslikke teadmisi, tuleks seda ka manuaali tõlkimisel silmas pidada. Erialalisi termineid tuleks kasutada ainult samades kohtades, kus neid on kasutatud lähtetekstis ja hoiduma uute lisamisest. Osades kohtades oli vaja lisada seletusi. KoGlossi manuaali tõlkimisel tuginesin KoGlossi eestikeelsele koduleheküljel olevale informatsioonile.

Tõlkes on täielikult üle võetud kõik algteksti illustratiivsed saksakeelsed joonised. Vaatasin ka võrdluseks projekti ingliskeelset manuaali. Saksa keelde on tõlgitud programmi *AntConc* abimaterjal⁴, mida ma olen tõlkimise ajal korduvalt vaadanud. Loomulikult lähtusin selle projekti manuaali tõlkimisel ka oma juhendaja Terje Looguse nõuannetest, kes nagu sissejuhatuses juba mainitud, osales KoGlossi projekti koostamisel. Tõlkija jaoks võib olla eriti väärtuslik võimalus suhelda inimesega, kes on lähteteksti üks autoritest või mingil teisel moel selle koostamisega seotud. Sel juhul oleks võimalik pärida kriitilist informatsiooni. Selline võimalus ei pruugi mitte alati olla olemas. Enamus töös kasutatud allikatest on internetiallikad, nagu erinevad masintõlget käsitlevad artiklid ja blogid. Kasutatud on ka Google'i ja Microsofti ametlikke väljaandeid. Teave „klassikaliseset“ masintõlkest ja selle ajaloost pärineb artiklist „Kuidas masin tõlgib“, mille autoriteks on Heiki-Jaan Kaalep ja Mare Koit.

4 http://www.laurenceanthony.net/software/antconc/resources/help_AntConc310_german.pdf

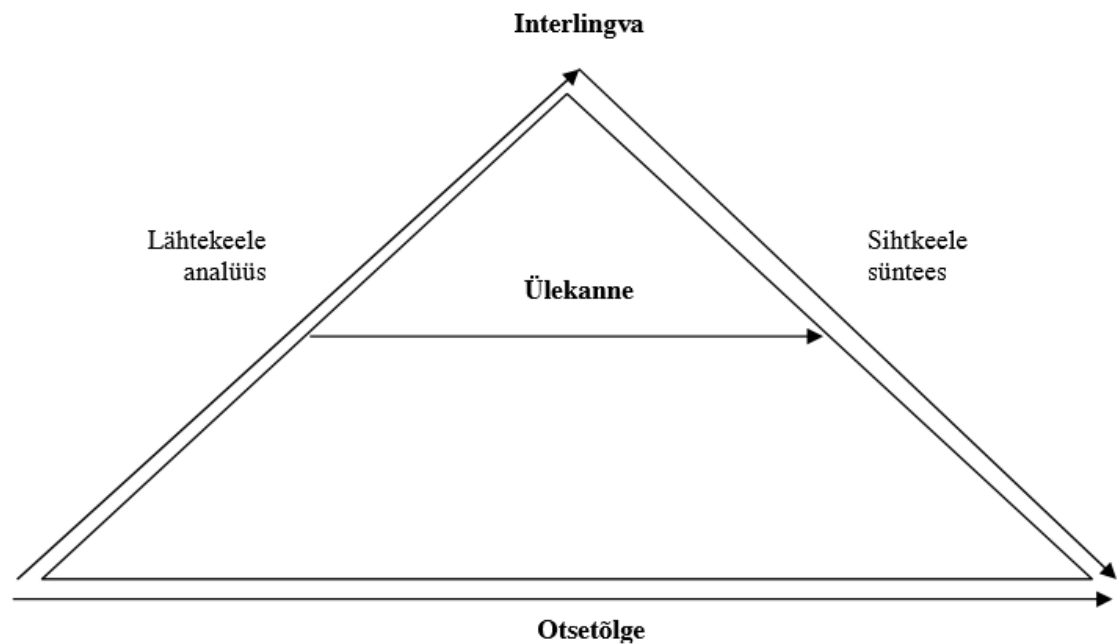
3. Masintõlkest

3.1. Masintõlke liigid

Sageli kiputakse arvama, et masintõlge on üks hiljutine nähtus, kuid see sai alguse juba esimeste arvutite loomisega, olles nende üks esimesi rakendusi. Masintõlke lähenemisviisideks on statistiline masintõlge ja reeglipõhine masintõlge, mida loetakse masintõlke „klassikaliseks lähenemiseks.“ Kõige hilisem masintõlke mudel on neurovõrgu põhine masintõlge. Reeglipõhine masintõlge (RPMT) jaguneb veel omakorda kolmeks tüübiks: otsesed süsteemid, ülekande RPMT süsteemid ja interlingual põhinevad RPMT süsteemid. Masintõlke klassikalised meetodid võttis kokku Bernard Vauquois (14. juuni 1929 kuni 30 september 1985) nn Vauquois' kolmnurgana. Selles nõuab iga kõrgem aste lähtekeele detailsemat analüüsi ja vastavalt sihtkeele detailsemat sünteesi. (vrd Kaalep, Koit 2010: 727)

3.1.1. Reeglipõhine masintõlge

Esimeseks liigiks on otsesed süsteemid, ehk sõnastikel põhinev masintõlge, mis seab sisendi ja väljundi vastavusse lihtsate reeglitega. Esimeste masintõlkesüsteemide puhul oli tegu otsetõlkesüsteemidega ja neid loodi tavaliselt ühe kindla keelepaari jaoks. Otsetõlke korral tõlgitakse lähtekeele tekst sõnahaaval, kasutades mahukat kakskeelset sõnastikku. Sisendit käsitletakse sõnavormide järjendina, millega sooritatakse mitmeid operatsioone. Lähtekeele sõnad asendatakse sihtkeele omadega ja muudetakse nende järjekorda saamaks korrektseid sihtkeelseid lauseid jne. Keeled peavad olema suhteliselt lähedase struktuuriga ja sarnase lauseehitusega, et otsetõlge oleks võimalik. (vrd Kaalep, Koit 2010: 727)



Vauquois' kolmnurk (Kaalep, Koit 2010: 727)

Teine lähenemisviis on ülekandel põhinev masintõlge, mis kasutab morfoloogilist ja süntaktilist analüüsi. Selle strateegia järgi jaguneb tõlkimine kolme faasi: lähtekeele analüüs, ülekanne ja sihtkeele süntees ehk genereerimine. Analüüsi faas jaguneb omakorda morfoloogiliseks, süntaktiliseks ja semantiliseks analüüsiks. Süntees jaguneb sama moodi semantiliseks, süntaktiliseks ja morfoloogiliseks sünteesiks. (vrd Kaalep, Koit 2010: 727)

Lõpuks on abstraktset keelteülest tähenduse esitamise viisi kasutavad interlingual põhinevad reeglipõhise masintõlke süsteemid. Selle käigus tõlgitakse tõlgitav tekst esmalt vahekeelde ja seejärel sihtkeelde. Mida rohkem keeli on vahekeelelega liidestatud, seda võimekam on selline tõlkesüsteem. (vrd Kaalep, Koit 2010: 728)

3.1.2. Statistiline masintõlge

Järgnevalt käsitletakse statistilise masintõlke mudelit, mida kasutavad Google ja Bing. Statistiline masintõlge kasutab kakskeelseid lähtekorpuseid, kus sama tekst on esitatud mõlemas keeles. Lähtekorpused, mida programm läbi töötab, koosnevad suurest hulgast tõlgitud tekstidest, näiteks raamatud ja veebilehed. Nende tekstide põhjal koostatakse statistiline mudel, mida saab kasutada tundmatute tekstide tõlkimiseks. Selle protsessi pideva kordamise tulemusel on lõpuks koos suur hulk mustreid, mille tulemusel õpib süsteem statistilise masintõlke korral ise tõlkima, ilma et keeleteadlane või programmeerija seda eraldi selgitama peaks. (vrd Fišel 2013: 1; Vikipeedia; Koitmäe, Mändul 2013: 52)

Kõige tuntumad statistilisel mudelil põhinevad süsteemid on Google ja Bing, kuid peale nende on olemas ka Yandex.Translate,⁵ mis võimaldab 2017. aasta juuni seisuga tõlkida 94 keele vahel. Statistilisel mudelil põhinevad näiteks ka QT21⁶ ja SUMAT.⁷ (vrd Fišel 2013: 1) Paari kursuste raames on külastanud Tilde grupi⁸ esindajad, kes tutvustasid oma ettevõtte arendatud statistilist masintõlketööriista. Näidati statistikat, mille kohaselt pidavat nende masintõlketööriist olema parem kui Google'i tõlge.⁹ Kuid paraku puudub selle magistritöö kirjutamise ajal tõlge saksa-eesti suunal. Juhul kui lähtekeel oleks olnud inglise keel, siis oleksin Google'i tõlke asemel hoopis Tilde masintõlketööriista kasutanud ja võrrelnud seda esimesega.

Statistilisel masintõlkel on aga üks puudus. Nimelt ei pruugi iga kogutud muster olla korrektne. Juhul, kui mingi ebakorrektnel muster hakkab liiga tihti korduma, tunnistab süsteem seda korrektse tõlkena. (vrd Bramowicz 2014) See puudutab küll Google'i tõlget, kuid teisi statistilisel mudelil põhinevatel süsteemidel võib sama probleem tekkida.

5 Yandex.Translate: <https://translate.yandex.com/> (18.06.2017)

6 QT21: Quality Translation 21: <http://www.qt21.eu/> (18.06.2017)

7 CORDIS Archive: http://cordis.europa.eu/fp7/ict/language-technologies/project-sumat_en.html (18.06.2017) (18.06.2017) (18.06.2017)

8 Tilde masintõlketööriist: <http://www.tilde.ee/> (18.06.2017)

9 <http://www.tilde.ee/vordlev-hindamine-naitab-et-tilde-letsmt-toimib-paremini-kui-google-translate> (18.06.2017)(18.06.2017)

3.1.2.1. Google'i tõlge

Koglossi manuaali tõlkimise käigus oli kõige olulisemaks kasutatud abivahendiks maailmas enimkasutatav statistiline masintõlketööriist *Google Translate* ehk Google'i tõlge¹⁰¹¹. Järgnevalt tutvustatakse selle erinevaid võimalusi. 2013. aasta mai seisuga kasutab seda 200 miljoni inimest päevas. Josh Estelle sõnul pärinevad 92 protsenti selle kasutajatest väljaspoolt Ameerika Ühendriike. (vrd Shankland 2013) See tasuta masintõlke teenus tõlgib 2017. aasta mai seisuga 103 keele vahel¹² Sinna on võimalik sisestada maksimaalselt 5000 tähemärki.

See populaarne vabavaraline masintõlketööriist, mida on võimalik kasutada peale brauseri ka mobiilsete brauserite, Androidi rakenduste, või iOS-i rakenduste kaudu, on ühtlasi ka töö üheks uurimisobjektiks. Nii brauseri kui ka mobiilseadme brauseri versioonid võimaldavad tõlkida tekste ja veebilehti ning mitte-mobiili brauser suudab tõlkida mõningaid dokumente. Android ja iOS-i rakendus saavad tõlkida teksti, reaajas kõnet, kujutisi, veebilehti ja mõnede keelte puhul isegi reaajas videosi. Rakenduste arendajatel on võimalik kasutada *Cloud Translation API*-d¹³ (*Application programming interface*), et arendada oma tõlketööriistu. (vrd Allan 2016)

Google'i tõlke lehekülje paigutus on suhteliselt lihtne. Leheküljel on kaks akent: vasakpoolse akna kohal olevas kastis tuleb valida lähtekeel ja aknasse kirjutada tekst või veebilehe aadress, mida soovitakse tõlkida, ja teise akna kohal olevas kastis tuleb valida sihtkeel. Kui on olemas Google'i kasutajakonto saab tõlgitud sõnad salvestada oma vestmikusse. Google'i tõlge näitab ka tõlgitud sõna liiki, nt tegusõna, nimisõna. Enamasti pakub see ka alternatiivseid sihtkeelseid vasteid. (vrd Vikipeedia - Google Translate)

Nagu juba öeldud kasutab Google'i tõlge statistilist masintõlget, mille käigus analüüsitakse miljoneid veebis olemasolevaid tõlgitud dokumente, et õppida sõnavara ja

10 Google'i tõlge: <https://translate.google.ee/?hl=et> (07.06.2017)

11 Inside Google Translate (YouTube video): https://www.youtube.com/watch?v=_GdSC1Z1Kzs (07.06.2017)

12 Languages - Google Translate: http://translate.google.com/about/intl/en_ALL/languages/ (07.06.2017)

13 *Cloud Translation API*: <https://cloud.google.com/translate/> (07.06.2017)

otsida keeles mustreid. (vrd Koitmäe, Mändul 2013: 53) Iga uue tõlgitava teksti puhul rakendab Google'i tõlge statistiliselt kõige tõenäolisemat mustrit, mis aga ei pruugi olla alati kõige korrektsemad. Nimelt lisab Google'i tõlge näiliselt korrektsed tõlked oma tõlkemälu hulka. Kui ebakorrektsete tõlgete arv kasvab liiga suureks, siis võib süsteem hakata lugema neid korrektsete tõlgetena. Selle tõttu võib süsteem muutuda hoopis potentsiaalselt „rumalamaks“. Juhul kui need ebakorrektsed tõlked välja arvata on programm siiski tänu sellele võimeline „õppima.“ (vrd Bramowicz 2014) Peter Norvigi sõnul ei ole see praegu veel tõsine probleem, kui see võib süveneda. Norvigi kohaselt hinnatakse mingi tõlgitud lehekülje tõlke kvaliteeti. Kui leitakse, et mingi tõlge ei ole piisavalt hea ei salvestata seda tõlkemälu hulka. (vrd Clark 2014)

3.1.2.2. Bingi tõlge

Bing Translator’i ehk Bingi tõlge¹⁴ on vabavaraline statistiline masintõlge, mis põhineb pilveteenusel Microsoft *Translator*¹⁵. 2017. aasta mai seisuga võimaldab see tõlkida 60 keele vahel. Seda saab kasutada brauseris või mobiilseadme brauseris. Pakutakse rakendusi järgmistele platvormidele: Windows, Windows Phone, iOS, Android, Apple Watch ja Android Wear. Seda on võimalik integreerida ka muude Microsofti rakendustega nagu näiteks Microsoft Office, Skype ja Visual Studio. Brauseri versioonid võimaldavad tõlkida ainult teksti ja veebilehti, kuid Microsoft Translator’i rakendused võimaldavad tõlkida ka reaajas kõnet, teksti ja pilte. (vrd Allan 2016)

Bing on oma paigutuse poolest suhteliselt sarnane Google’i tõlkega. Leheküljel asuvad samuti kaks akent, vasakpoolsesse aknasse tuleb sisestada lähtekeelne tekst või veebilehe aadress ja parempoolsesse aknasse ilmub selle sihtkeelne versioon. Mõlema akna kohal on kast, milles tuleb valida lähtekeel ja sihtkeel. Akende keskel asub nupp „Tõlgi“ või „Translate.“ Selle kohal asub hall ümmargune nupp, mis vahetab lähtekeele ja sihtkeele vahel. Bingi üks erinevusi on see, et tõlke peale klõpsates ei pakuta võimalike sihtkeelseid alternatiive.

Oma tõlkeseminaride raames tehtud tõlketööde kirjutamisel olen seda peale Google’i tõlke paaril korral kasutanud. Vähemalt nende kogemuste põhjal väidaksin, et Google’i tõlge on parem. See võib tähendada ka seda, et tõlge saksa-eesti suunal ei pruugi olla veel samal tasemel kui mingil teisel suunal. KoGlossi manuaali tõlkimisel mina Bingi tõlget ei kasutanud, kuid analüüsis võrreldake kuivõrd kasulik on Bingi tõlge valitud katkendite tõlkimisel.

14 Bingi tõlge: <http://www.bing.com/translator?from=&to=et&setlang=et> (07.06.2017)

15 Microsoft Translator: <https://translator.microsoft.com/> (07.06.2017)

3.1.3. Tehisnärvivõrgu põhine masintõlge

Neurovõrgu põhine masintõlge kasutab tehisnärvivõrke. See on suhteliselt hiljutine masintõlke mudel. 2016. aasta lõpus tuli Google välja GNMT-ga (*Google Neural Machine Translation*). Google'il on plaanis vahetada senine statistiline masintõlge GNMT vastu välja. Sarnaselt statistilisele mudelile esineb probleeme keelte puhul, mis on omavahel väga sarnased. (vrd Sibold 2016) „Näiteks tõlkides lauseid inglise keelest, kas prantsuse või hispaania keelde. Samuti tuleb endiselt ette juhuseid, kus tõlkeprogramm üritab tõlkida nimesid või termineid. Juhtub ka seda, et tõlkides kaovad osad sõnad või muutub lause sisu.“ (Sibold 2016) Selliseid vigu on teinud ka statistilisel mudelil põhinevad masintõlketööristad Google ja Bing KoGlossi manuaali tõlkimisel. Google'i mobiilseadme rakendused kasutavad GNMT-d terves ulatuses, mis teeb 18 miljonit tõlget päevas. (vrd V. Le, Schuster 2016)

Neurovõrgu põhine masintõlget on hakanud kasutama ka Microsoft, nimelt mobiilseadme rakendused Skype Translator ja Microsoft Translator. Tartu Ülikooli arvutiteaduse instituudi keeletehnoloogia uurimisrühm on loonud oma neurovõrgu põhise tarkvara,¹⁶ mis võimaldab käesoleva magistritöö koostamise ajal tõlkida eesti ja inglise keele vahel. Harvard NLP grupp on avaldanud avatud lähtekoodiga neurovõrgu põhise masintõlketöörista OpenNMT¹⁷. (vrd Vikipeedia) Tilde on ka tulnud välja oma neurovõrgu põhise masintõlketööristaga Neural MT.¹⁸

16 Tartu Ülikooli masintõlge: <http://neurotolge.ee/> (07.06.2017)

17 *An open-source neural machine translation system*: <http://opennmt.net/> (07.06.2017)

18 Tilde Neural MT: <https://translate.tilde.com/neural/et/> (17.06.2017)

3.2. Masintõlke võimalused ja kasulikkus tõlkimisel

Nagu sissejuhatuses juba mainitud, on käesoleva magistritöö eesmärgiks välja selgitada, millisel tasemel on koostamise ajal kättesaadavad masintõlketööriistad ja kui palju oleks nendest abi tõlkeülesannete lahendamisel. Vaadatakse ka milline eelnevalt tutvustatud kahest kõige populaarsemast masintõlketööriistast on käesoleva magistritöö koostamise ajal kõige efektiivsem.

Inimtõlget mõjutavad nii lähtekeele kui ka sihtkeele omadused, kultuurikontekst ja üksikute tõlkijate tõlkimise võimekus. Kuna erinevad tõlkijad mõistavad lähtekeelt ja sihtkeelt erinevalt võivad nende tõlked varieeruda. Seevastu masintõlke puhul jääks individuaalne tõlkimise võimekus ära. (vrd Li, Graesser, Cai 2014: 191) Seejuures on oluline mõista, et masintõlge ei ole täiuslik. Masin võib tõlkida ära lause, kuid mitte mõtet. (vrd Goodman 2016) Seega peab ikkagi inimene masina teksti üle vaatama ja vajadusel korrigeerima. Selle küsitava tõlke kvaliteedi tõttu kipub masintõlkel olema üldiselt suhteliselt negatiivne maine. Seega võib tekkida küsimus, et miks üldse kasutada masintõlget. Näiteks tavakasutaja seisukohast on masintõlke peamisteks eelisteks, et see on tasuta ja kiire. Juhul kui kellegi eesmärgiks on saada tekstist kätte ainult kõige põhilisema informatsiooni ja tõlge jääb isiklikuks tarbeks, ei oma ortograafia ja vormistus tähtsust. Ettevõtja seisukohalt võib masintõlge olla samuti mõistlikum kui inimtõlge, sest inimtõlge on võrreldes masintõlkega palju kulukam ja ettevõtte eelarve on piiratud. Ettevõttel võib olla vaja tõlkida näiteks veebileht, kus on ainult üksikud sõnad (nt *menu*, *about*, *faq*), mida ei ole raske vajadusel korrigeerida. Masintõlget kasutatakse laialdaselt standartsete tekstide tõlkimisel, nt ilmateated. Masintõlget kasutatakse aktiivselt veebiteenustena ettevõtetes nagu näiteks Autodesk ja Euroopa Liidu poolt rahastatakse mitmeid masintõlke projekte nagu näiteks QT21 ja SUMAT. (vrd Fišel 2013: 1)

Käesoleva magistritöö koostamise ajal võimaldab Google'i tõlge tõlkida 103 keele vahel ja Bingi tõlge 60 vahel. Masintõlketööriistu aga ei ole mõistlik hinnata üksnes selle järgi, kui mitme keele vahel need tõlkida võimaldavad. Nimelt ei oma iga keel samasugusel tasemel toetust ja seega ei ole võimalik kasutada igat pakutud funktsiooni

mistahes keelega. Näiteks Prantsuse keele puhul toimivad kõik kuus funktsiooni, milleks on trükkimine (*type*), käega kirjutamine (*write*), rääkimine (*talk*), pildid (*snap*), tõlge telefoni kaameraga (*see*) ja võrguta režiim (*offline*). (vrd Allan 2016) Eesti keelega toimivad kõik funktsioonid peale *snap* ja *see*.¹⁹

Microsoft *Translator*'i puhul on olukord sarnane, kuid palju väiksemas ulatuses. Reaalajas kõne tõlkimine toimib ainult järgmiste keeltega: araabia, mandariini hiina, inglise, prantsuse, saksa, itaalia, portugali, vene ja hispaania. Kuid see-eest on võimalik kasutada 60-st keelest 44 keelt *offline* režiimis, Google'i tõlge toetab ainult umbes poolte *offline* režiimis kasutamiseks. Täielku tõlketoe pakkumisel peamiste keelte osas on need samal tasemel. (vrd Allan 2016)

Mõlemad masintõlketööriistad põhinevad statistilisel mudelil. See kuivõrd suutlikud need on mingit teksti sihtkeelde panna, määrab ära tegelikult tekstide hulk. Nimelt kasvab tõenäosus, et soovitud tõlkega sarnaseid lausemalle võib leidub mõnest masintõlketööriistale aluseks olevast tekstist, iga uue tekstiga. Selle pärast on paljude keelte tõlkekvaliteet kehvem. Tõlkekvaliteeti võivad kindlasti mõjutada ka serveri maht ja kiirus, sest teksti sisestamisel on programmil vaja koheselt läbi töötada miljardeid tekste ja nende hulgast kõige tõenäolisemad mustrid välja valima. (vrd Koitmäe, Mändul 2013: 53) Tõlkekvaliteet oleneb peale tõlgitava teksti ka tõlkesuusnast ning mida ja kui palju sinna sisestada. Tulemused võivad vahest suurel määral varieeruda, kas sisestada üksikuid sõnu, lõike või tervet teksti. Juhul kui masintõlge on suuteline pakkuma piisavalt heal tasemel väljundit, siis võib tõlketöö järeltoimetamise korras sujuda kiiremini kui nullist tõlkimisel. (vrd Fišel 2013: 1)

Järgnev näide ei ole küll otseselt minu magistritööga seotud ja ei pärine sellest, kuid võib tunduda huvitavana ja illustreerida, kuidas Google'i tõlge võib mingite sõnade puhul toimida. Kuna mina ei oska eriti hästi vene keelt, kuid soovin siiski mingi mind huvitava teema kohta rohkem teada saada, siis võib öelda olen põhimõtteliselt sõltuv masintõlkest. Kasutan seda ka muude võõrkeelsete lehekülgede puhul. Kindlasti leidub

19 http://translate.google.com/about/intl/en_ALL/languages/ (02.05.2017)

hulgaliselt inimesi üle maailma, kes on sarnases olukorras. Kui sisestada Google'i tõlkesse näiteks inglise „*wiper arm*“ (klaasipühkija vars), annab Google'i tõlge vasteks vene „*рычаг (стеклоочистителя)*“. See vaste ei ole sugugi väär, kuid vähemalt mina olen tähele pannud, et palju sagedamini kasutatakse hoopis sõna „*новодок*“ kui „*рычаг*“. Siiski kasutatakse mõlemaid sõnu nii ametlikult²⁰ kui ka tavainimeste poolt. Selleks et Google'i tõlge annaks sagedasema „*новодок*“ tuleb sinna sisestada inglise „*leash*“. Võib arvata inglise-vene suunal on tunduvalt rohkem kasutajaid kui inglise-eesti suunal ning seega peaks pakutavate vastete hulka ilmuma kõige sagedasemad. Google'i tõlkel on samuti vahetevahel probleeme lähtekeelega tuvastamisega (nt vene ja ukraina).

20 „*рычаг*“: <http://www.avtopribor.ru/index.php/ru/production.html?page=shop.browse> (02.05.2017)

21 „*новодок*“: http://www.herzog-germany.ru/download/Herzog_Germany_2011.pdf (02.05.2017)

4. Tõlkeanalüüs

4.1. BLEU

Võib tekkida küsimus, kuidas või mille järgi saaks hinnata masina poolt tõlgitud teksti? Selle tarbeks on loodud algoritm BLEU (*Bilingual Evaluation Understudy*), mille abil on võimalik hinnata mingi teksti kvaliteeti, mis on masina poolt tõlgitud ühest loomulikust keelest teise võrreldes inimtõlgitud tekstiga, mille on tõlkinud professionaalne tõlkija. BLEU oli üks esimesi meetrikaid, mis suutis saavutada kõrge korrelatsiooni inimhindajate poolt tõlgetele antud punktidega. See on siiaeni jäänud populaarseimaks automatiseeritud meetrikaks, mis ei ole kulukas. (vrd Papineni, Roukos, Ward, Zhu 2002; Kirik 2008; Vikipeedia)

BLEU meetrika põhimõte seisneb selles, et mida lähedasem on mingi masintõlketööriista genereeritud tõlge professionaalse inimtõlkija loodud tõlkele, seda võimekamaks võib antud masintõlketööriista tõenäoliselt pidada. Kuna BLEU meetrika hindab ainult masintõlke sarnasust etaloniks antud inimtõlkega ei arvestata võimalikke alternatiivseid variante, mis ei pruugi olla väärad. BLEU meetrika ei arvesta ka otseselt tõlke grammatilist või sisulist korrektsust. (vrd Kirik 2008: 7; Vikipeedia) Käesoleva magistritöö analüüsis on etaloniks minu tehtud tõlge. Töö analüüsi peatükis toodud näidete BLEU hinnangute arvutamiseks on kasutatud Tilde Interactive BLEU score evaluator'it.²²

22 Tilde Interactive BLEU: <https://www.letsmt.eu/Bleu.aspx> (19.06.2017)

4.2. Masintõlke võimekus keskse terminoloogia tõlkimisel

Selles alapeatükis vaadatakse töö keskset terminoloogiat ja kuivõrd võimekas on masintõlge selle tõlkimisel. KoGlossi nime puhul on tegu lühendiga, mis koosneb järgmiste sõnade elementidest: **k**orpuspõhine, **k**ollaboratiivne, **k**onstruksioonide, **g**lossarium. Google andis „*korpusbasierte*“ vasteks „korpuspõhise“ ja Bing „korpuse baasil“, „*kollaborative*“ vasteks „koostöö“, mis niisama ei ole sobilik, kuid sellest oleks võimalik tuletada „koostööpõhine.“ Antud vaste sobiks hästi KoGlossi nimesse, kuna selles esineb element „ko“. See oleks võibolla kõige tõenäolisem vaste, mille ma oleks valinud juhul, kui KoGlossi projektil puuduks eestikeelne kodulehekülg. Pakutud vastete hulgas oli antud ka „koostööprojekte“, mida KoGloss tõepoolest on.

Algtekstis ja sihttekstis on üheks tähtsaks keskses terminiks (*das*) *Glossar* ehk glossarium, mis on ühtlasi osa meetodi nimest. Sõnaraamatus on antud sellele järgnev vaste:

„*Glossar n.1c glossaar, seletussõnastik*“ (Kibbermann, Kirotar, Koppel 2007: 496)

Masintõlketööriistad andsid tekstis „(*das*) *Glossar*“ vasteks kohati nii „sõnastik“ kui ka „sõnaraamat“, mis ei ole samuti põhimõtteliselt väärad, kuid käesolevas kontekstis ei ole need kõige sobilikumad. Tekstis õpetatakse, kuidas seda on võimalik ja mis eesmärgil koostada. Otsustasin glossariumi kasuks, sest minu tõlke üheks eesmärgiks oli olla kooskõlas olemasoleva tekstiga KoGlossi koduleheküljel. See termin võib tunduda enamusele võõras.

Üheks teiseks keskses terminiks on konstruksioon ehk (*die*) *Konstruktion*, mis on samuti pealkirja sees. Sõnaraamatus on antud järgnev vaste:

„*Konstruktion f.1b Konstruksioon, tarind*“ (Kibbermann, Kirotar, Koppel 2007: 496)

Need on keelelised konstruksioonid, mida hakatakse ise koostatud digitaalsetest (teksti)korpustest, mis on samuti keskne termin, otsima ja glossariumisse kandma. Need korpused koosnevad erialakeelsetest tekstidest.

Selles töös on üheks keskseks terminiks masintõlketööriist. Ettevõtte Tilde kasutab seda oma koduleheküljel loodud masintõlke programmi kirjeldamiseks. Minu arvates väljendab see mõiste väga hästi käesoleva magistritöö kirjutamise ajal saadaval oleva masintõlke olemust ja taset ning kasutasin seda läbivalt oma töös. Mark Fišel kirjutas ka oma artikklis „Milline on hea masintõlge?“, et masintõlget „...tuleb võtta tööriistana, mis võib vähendada inimtõlkija tööruutiini.“ (Fišel 2013: 1) Töö analüüsi käigus selgus, et masintõlge võib olla tõepoolest kõigest abivahend, kuid mitte iga teksti puhul. Selle alla käivad klassikaline, statistiline ja esialgu ka tehisenärvivõrgu põhine masintõlge.

4.3. Masintõlke võimekus KoGlossi manuaali tõlkimisel

Selles alapeatükis selgitatakse välja kuivõrd võimekas on masintõlge ja seda kui kasulik on see KoGlossi manuaali tõlkimisel. Nagu eelmises peatükis öeldud ei pakuta igale keelele samasugusel tasemel toetust. Tavaliselt saavad suuremad keeled uuendused esimesena ning kasutajate ja tõlgitud tekstide arv on suurem, mis on statistilise mudeli puhul oluline. Tõlke kvaliteeti mõjutab kindlasti ka kuivõrd lähedased on mingid keelepaarid omavahel. Teatavasti kuulub Saksa keel Germaani keelterühma, mis omakorda Indoeuroopa keelkonda. Eesti keel kuulub aga Soome-ugri keelte hulka, mis omakorda Uurali keelkonda. Alapeatüki teises pooles käsitletakse üldiseid masintõlkega seonduvaid probleeme, tähelepanekuid ja muid erinevaid iseärasusi. Selles alapeatükis vaadatakse peamiselt Google'i tõlke võimekust, kuna seda on kasutatud manuaali tõlkimisel. Bingi tõlget käsitletakse järgnevas alapeatükis.

Töö käigus selgus, et Google'i tõlke üheks läbivaks probleemiks on sellised saksakeelsed sõnad nagu „*Die NutzerInnen*“ (kasutajad) või „*Die Teilnehmer/innen*“ (osalejad). Need sõnad esinevad Moodle'it käsitlevas peatükis. Nimelt ei ole programm suuteline tuvastama seda, et need sõnad hõlmavad nii meessoost kui ka naissoost inimesi. Võiks eeldada, et need vormid peaksid juba ammu olema Google'i tõlkemälus ja tõlgib need kohe vastavalt „kasutajad“ ja „osalejad“. Kui sisestada Google'i tõlkesse „*Die Teilnehmer/innen*“ anntakse vasteks „Osalejad / interjäär“. Siit selgub, et programmi jaoks on „*innen*“ eraldi sõna. Vasteks võidakse anda ka näiteks „sees“, „seestpoolt“ või isegi inglise „*inside*“. Kuna „*innen*“ üksi tähendab saksa keeles tõepoolest „sees“, siis sel juhul ei ole tegu väärast tõlkega. Kui valida sihtkeeleks inglise keele toimib programm samamoodi, andes vasteks „*Participants / inside*“. Bing seevastu andis kohe vasteks „osalejad“. Samuti kohtleb programm sõnade „*Lehrende/r*“ ja „*Trainer/in*“ lõpus kaldkriipsuga eraldatud „r“ ja „in“ eraldi olevate üksustena, andes vasteteks „õpetajad / r“ ja „Trainer / in“.

KoGlossi meetodi potentsiaalsete kasutajate hulka kuuluvad ka „*PR(Public Relations)-Leute*“ ehk suhtekorraldajad. Sõna esineb peatükkides üks ja kuus ning „*PR-Leute*“ puhul käitus Google'i tõlge sarnaselt nagu „*Lehrende/r*“ ja „*Trainer/in*“ puhul ehk

programm kohtleb lühendit „PR“ eraldi. Täpsemalt lühend „PR“ ei ole programmi jaoks lühend, vaid tõlgitakse otse „PR-inimesed“. Kui valida sihtkeeleks inglise keele, tuvastab programm lühendi ära ja pakutava vastena ilmub „*Public relations*“. Tõenäoliselt võib põhjus olla selles, et mõlemas keeles kasutatakse seda lühendit. Sellele on eestikeelne vaste leitud Google'i otsingumootori kaudu.

Algtekstis *Antconci* kasutamist käsitlevas peatükis esineb tihti sõna „*Reiter*“, millega käesolevas kontekstis peetakse silmas keeletarkvaras *Antconc* olevaid kaarte (*Concordance, Concordance Plot, File View, Clusters, Collocates, Word List, Keyword List*). Esimene vaste, mida Google'i tõlge sellele annab on „sõitja“ mitte „ratsanik“. „*Der Reiter*“ võib tõepoolest tähendada ratsaniku. Üheks vasteks on pakutud ka „Reiter'i“ ehk Reiter kui kellegi/millegi nimi.

Järgnevalt võrreldakse kuivõrd kasulikud on masintõlketööriistad *Antconci* funktsioone seletavate ridade tõlkimisel. Tekstis leidub hulgaliselt selliseid ridu, kus on inglise keeles mingi funktsiooni nimi ja seejärel selle saksakeelne selgitus. Masintõlketööriistade jaoks võib olla vägagi problemaatiline, kui ühes lauses on korraga kaks keelt. Seega võib langeda tõlke kvaliteet ja suureneb järeltoimetamise vajadus. Selguse huvides on masintõlketööriistade poolt tõlgitud ridade esimesed pooled rasvases kirjas nagu algtekstis ja toimetatud tõlkes.

a) Originaal: „***Treat all Data as Lowercase: Ist diese Einstellung aktiviert, wird nicht zwischen Groß-und Kleinschreibung unterschieden.***“ (KoGloss 2012: 17)

b) Google'i poolt tõlgitud toimetamata rida: „**Kohtle andmeid Väiksed tähed:** Kui on lubatud, ei ole tõstutundlik.“

c) Bingi poolt tõlgitud toimetamata rida: „**Kõik andmed käsitlema väiketähed:** see säte on lubatud, juhul ei Erista suurtähti ja“

d) Toimetatud rida: „***Treat all Data as Lowercase:*** Kui see seadistus on aktiivne ei tehta vahet suur- ja väiketähedel.“

Nende ridade põhjal oleks võimalik väita, et selliste teksti kohtade juures ei ole tõlkijale ei Google'i ega ka Bingi tõlkest erilist abi. Toimetemata tõlgete sisust ei ole võimalik aru saada. Kuigi Google'i tõlge pakub sõnale „tõstutundlik“ ühe võimaliku alternatiivina „suur- ja väiketähtede tundlikud“, mis oleks korrektsem kui masintõlketööriista esimene pakutud vaste. Mõlemad masintõlketööriistad tõlkisid ära ingliskeelse funktsiooni nimetuse „*Treat all Data as Lowercase*“, mida aga ei ole vaja. Kuid seda ei ole võimalik seadistada, mida tõlkida ja mida mitte. Tuleb sisestada ainult osa, mida soovitakse tõlkida. Kusjuures esimese näitena toodud toimetamata katkendis on ingliskeelne osa - „*Center for English Language Education in Science and Engineering*“ - jäetud nii nagu algtekstis. Võiks arvata, et suhteliselt levinud saksakeelne konstruktsioon „*wird nicht [...] unterschieden*“ peaks olema programmile juba tuttav, kuid seda ei suudetud tuvastada. Probleemide põhjus võiks peituda näiteks selles, et tõlgitavas reas esinevad inglise ja saksa keel koos. Kui sisestada „*wird nicht unterschieden*“ eraldi kujul Google'i tõlkesse annab programm korrektse vaste „Ei tehta vahet.“ Tõlkijal oleks seega vaja terve punkt otsast peale ümber kirjutada. Sellelaadsete ridade puhul oleks mõtekam kopeerida ingliskeelne funktsiooni nimetus ja kirjutada selle selgitus täielikult ise.

Käesoleva töö kirjutamise ajal saadaval olevad masintõlketööriistad on tõepoolest suutelised tõlkima lihtsamaid laused suhteliselt heal tasemel. Vajades seejuures ainult paaris kohas toimetamist. Fišel tõi selle kohta hea näite, kus ingliskeelse „do not buy their products“ eestikeelne tõlge võib olla näiteks „ostke nende toodangut“. Selle tõlke sisu on lähtelausega vastupidine, kuid selle parandamiseks oleks vaja lisada lause algusse vaid sõna „ärge“. (vrd Fišel 2013: 1) Tööst pärinev sellele sarnane näide oleks „näiteks suheldes välismaiste äripartnerite“, kus tuleb lisada ainult viimasele sõnale liide „-ga“. Järgnevas alapeatükis selgub, et sellise teksti nagu Koglossi manuaali puhul, oleks vaja tõlke toimetajal teha palju enam kui lisada mõned sõnad või liited.

4.4. Masintõlketööriistade kasulikkus nelja näite põhjal

Selles alapeatükis võrreldakse vabavaraliste masintõlketööriistade Google *Translate* ja Bing *Translator* tõlkevaliteeti ja kasulikkust, nelja näite põhjal. Esmalt on antud võrdluseks vastavate näidete originaalid ja masintõlketööriistade poolt genereeritud tõlked. Nende masintõlketööriistade poolt genereeritud tõlgete järel analüüsitakse nendes leiduvaid erinevaid vigu ja muid iseärasusi. Arutletakse millises ulatuses oleks vaja neid toimetada. Arvestatud on ka seda, et magistritöö kirjutamise ajal ei ole masintõlketööriistad veel suutelised pakkuma suhteliselt rahuldava tõlget. Lõpus on antud ka kolme viimase vaatluse all oleva näite BLEU hinded, mis on arvutatud kasutades Tilde Interactive BLEU score evaluator-it. Kuna magistritöö üheks eesmärgis uurida, kas masintõlge aitab aega kokku hoida või mitte, on viimased kolm näidetena toodud lõigud sisestatud mõlemasse masintõlketööriista tervelt. Tõlke tulemust võib erineda, kas sisestada üksikute lausete või lõikude kaupa.

4.4.1. Näide 1.

„[...] *Beim Lehren und Lernen rückt man mittlerweile vom Pauken isolierter Vokabeln ab und greift auf die Vermittlung miteinander auftauchender Wörter und sprachlicher Konstruktionen zurück.* [...]“ (KoGloss 2012: 4)

a) Google'i tõlgitud näide.

„Õpetamise ja õppimise üks liigub nüüd timpanid isoleeritud sõnavara ja kasutab vahenduse areneva koos sõnade ja keelelise struktuure.“ (10.04.2017)

b) Bingi tõlgitud näide.

„Õpetamise ja õppimise üks liigub eemale timpanid isoleeritud sõnavara ning põhineb vahendusega seotud sõnad ja keel.“ (10.04.2017)

Selle lause põhjal on näha, kuidas mõlemad statistilisel mudelil põhinevad masintõlketööriistad ei arvesta tõlkimisel konteksti. Mõlemad tõlkisid tegusõna

„*pauken*“ timpaniks. Selles kontekstis aga ei ole mitte tegu löökpilliga *Die Pauke* vaid kõnekeelse tegusõnaga *pauken* ehk tuupima. Samas lauses esinevad ka õpetama ja õppima, inimtõlkija oleks suuteline selle koheselt ära tabama, kuid tööriistad mitte. Tõlgitud lausete sarnasuse põhjal võiks oletada, et mõlema tööriista tõlkemälus ei leidu piisavalt tekste, kus see kõnekeelne tegusõna esineb. Masintõlketööriista poolt tõlgitud teksti toimetamisel tuleks arvestada sellega, et antud vasted ei pruugi olla alati kõige sobilikumad ja oleks mõttekas neid kontrollida. Bingi näites on „*sprachlicher Konstruktionen*“ „*Konstruktionen*“ osa ära kadunud, kuid Google'i omas on see alles. Selle näite põhjal saaks kahelda masintõlke kasutamise mõttekuse üle, kuna inimesest tõlkija arvestab vastete valikul ka konteksti. Tõlgetes esineb palju probleeme ka lausete sidususe ja grammatilise korrektsusega.

4.4.2. Näide 2.

„*AntConc wurde von Professor Laurence Anthony, dem Leiter des CELESE (Center for English Language Education in Science and Engineering) an der Universität Waseda in Tokyo entwickelt und ist besonders für kleinere Textsammlungen (bis zu 20 Mio. Textwörter) geeignet. Es ist ein viel genutztes Programm zur Analyse digitaler Textsammlungen, das einfach zu bedienen ist, nicht installiert werden muss, und zu dem eine ausführliche Dokumentation und weitere Ressourcen verfügbar sind (siehe: http://www.antlab.sci.waseda.ac.jp/software/README_antconc3.1_german.pdf). Die verfügbaren Tools erlauben Anfragen zum Erstellen von Wortlisten, Clustern, Konkordanzen, Kollokationen und eine detaillierte Einsicht in die Kontexte der Suchergebnisse.*“ (KoGloss 2012: 9)

a) Google'i tõlgitud näide 2. (12.05.2017)

„AntConc oli professor Laurence Anthony juht Celese (Center for English Language Education in Science and Engineering) arendada Tokyo Waseda Ülikooli ning on eriti sobiv väikestele tekstikogude (kuni 20 miljonit teksti sõna) sobivad. See on laialdaselt kasutatud programmi analüüsisides digitaalse teksti kogud, mis on lihtne kasutada, ei vaja

installi ja ulatusliku dokumentatsiooni ja muude ressursside olemasolu (vt [http://www.antlab.sci.waseda.ac.jp / tarkvara / README_antconc3.1_german.pdf](http://www.antlab.sci.waseda.ac.jp/tarkvara/README_antconc3.1_german.pdf)). Olemasolevad vahendid võimaldavad taotluste luua sõna nimekirjad, klastrid, kooskõlastused, collocations ja üksikasjaliku ülevaate kontekstis otsingutulemustes.“

b) Bingi tõlgitud näide 2. (10.05.2016)

„AntConc oli Professor Laurence Anthony, CELESE (inglise keeles hariduse, teaduse ja inseneriteaduse vallas keskus) Waseda Ülikooli Tokyo juhataja ning on eriti sobilik väiksemate kogud tekstide (kuni 20 miljonit puhul). See on palju-kasutada programmi analüüsides digiteksti kogumeid, mis on lihtne kasutada, ei tohi paigaldada ja üksikasjalikud dokumendid ja täiendavate vahendite olemasolu (vt: http://www.antlab.sci.waseda.ac.jp/software/README_antconc3.1_german.pdf).

Kättesaadavad vahendid võimaldavad taotlusi luua klastrid, kooskõlastused, sõna nimekirjad, tähendust ja kasutusviisi ja üksikasjaliku ülevaate otsingutulemuste kontekstides.“

Selles lõigus leidub peale saksakeelse teksti ka mõningad inglisekeelsed sõnad. Google tõlkis ära ka lingi sees oleva „*software*“ (tarkvara), samas kui sulgudes olev ingliskeelne seletus lühendile *CELESE* „*Center for English Language Education in Science and Engineering*“ on jäetud Google'i poolt tõlkimata. Huvitav tähelepanek on see, et algtekstis on lühend „*CELESE*“ suurte tähtedega ja tõlkes enam mitte. Sellest võib järeldada, et programm kohtles seda lühendit nimena. Samas lingi sees olev „*README*“ on jäetud samaks. Siiski oli Google'i tõlge suuteline ära tõlkima kõige põhilisema informatsiooni.

Terve Bingi tõlgitud lõigu sisu on võrreldes Google'iga vaevu mõistetav. Selles lõigus on näha ka tähenduse moonutust. Keeletarkvara *Antconci* puhul on tegemist sellise tarkvaraga, mida ei ole tarvis arvutisse installeerida vastupidiselt näiteks Microsoft Office või Adobe Photoshop. See tähendab, et *Antconc* on kohe pärast allalaadimist kasutusvalmis. Google'i tõlge oli suuteline mingil määral algteksti mõtte ära tabada, et

Antconc „ei pea olema paigaldatud“. Seevastu Bingi tõlkest võib välja lugeda, et *Antconci* „ei tohi paigaldada.“ Mingil ajahetkel „AntConc oli Professor Laurence Anthony“ ise. Seda Bingi tõlgitud lõiku oleks vaja toimetada tunduvalt suuremas ulatuses kui Google'i tõlgitud lõiku. Seda kinnitavada ka vastavad BLEU hinded. Google'i tõlgitud lõigu hindeks on 41.93 sajast ja Bingi tõlgitud lõigu hindeks on 32.98.

4.4.3. Näide 3.

„Dieses Grundmuster kann durch Lehrende, Studierende oder Berufsnutzer nach Bedarf ergänzt, gekürzt oder modifiziert werden. Das Grundmuster enthält die Angaben zur Morphologie, Syntax, Semantik (Bedeutung) und zum Gebrauch der Konstruktion (Stil, Textsortenspezifik, Verwendungshäufigkeit usw.). Der Eintrag enthält auch Verwendungsbeispiele aus dem Korpus. Der Verfasser des Eintrags hat auch die Möglichkeit, wichtige Erkenntnisse aus der Korpusanalyse oder aus seinem sonstigen Wissen in den Glossareintrag zu schreiben, unabhängig davon, ob es sich um grammatische, semantische, pragmatische oder sachlich informative Aspekte handelt. Am Ende des Eintrags können der Verfassersname und das Datum angegeben werden.“
(KoGloss 2012: 25)

a) Google'i tõlgitud näide 3. (12.05.2017)

„See muster võib täiendada õpetajate, õpilaste või professionaalsete kasutajate vaja vähendada või muuta. Põhimudel sealhulgas üksikasjad morfoloogia, süntaks, semantika (täendus) ja kujunduse kasutamine (stiil, teksti tüüpi spetsiifilisus, kasutamise sagedus jne). Piiril ka kasutuse näiteid corpus. Autor kanne on ka võime kirjutada peamised järeldused corpus analüüsi või selle teiste teadmiste sõnastiku kanne, olenemata sellest, kas see on grammatiliselt, semantiline, pragmaatiline või faktiliselt informatiivne aspekte. Lõpus kande autori nime ja kuupäeva saab määrata.“

b) Bingi tõlgitud näide 3. (10.05.2016)

„Selles põhimudeli saab täiendada õpetajate, õpilaste või professionaalsete kasutajate nõudlust, vähendada või muuta. Viimase peamine mudel sisaldab teavet morfoloogia, süntaks, semantika (st) ja kasutada disaini (stiil, teksti tüüp spetsiifilisus, Tarbimissagedus, jne.). Kanne sisaldab keha kasutamise näited. Autor kanne on ka võime kirjutada, olenemata sellest, kas see on grammatilisi, semantiline, pragmaatiline või faktiliselt informatiivne aspektid olulised järeldused keha analüüsi või muud faktidesse sõnastik kanne. Kirje lõpus saate määrata nimi, Autor ja kuupäev.“

Mõlema masintõlketööriista jaoks võib olla keeruline mõista esimeses lauses esinevat saksakeelset modaalse tegusõnaga passiivi konstruktsiooni „[...] *kann durch Lehrende, [...] modifiziert werden*“, mis otsetõlkes tähendaks „saab läbi õpetajate muuta.“ Selle lausega on tegelikult mõeldud seda, et õpetajatel on võimalik seda põhimustrit muuta. Oma tõlkes kasutasin „saavad õpetajad, [...] muuta. Algtekstis on püütud „*können*“ asemel kaudse „*kann durch*“ kasutamisega olla ametlikum ja distantseeruda. Kui aga sisestada algupärase lause asemel hoopis „*Dieses Grundmuster können Lehrende, Studierende oder Berufsnutzer nach Bedarf ergänzen, kürzen oder modifizieren*“, on tulemuseks „See põhi muster võib õpetajad, õpilased või professionaalse täiendada, kustutada või muuta kasutajad nagu vaja,“ mis algsest toimetamata lausest eriti palju ei erine.

Selles lõigus on näha Google'i tõlke ühte läbivat iseärasust, nimelt saksa „*(das) Korpus*“ on tõlgitud inglise „*corpus*“-ks. Erinevalt Google'ist, mis tõlkis „*(das) Korpus*“ inglise „*corpuseks*“, tõlkis Bing selle „kehaks“. Kui sisestada „*das Korpus*“ üksinda, siis ilmub korrektne vaste „korpus“. Sulgudes olev „*(Bedeutung)*“ on tõlgitud „(st)“-ks. Kuna saksa „*(die) Bedeutung*“ on tõlkes tähendus, tuvastas programm seda kui „st - see tähendab.“ Google'i tõlgitud lõigu BLEU hindeks on 20.16 ja Bing on 11.78.

4.4.4. Näide 4.

*„Die Angaben zur Morphologie (Informationen über Wortart und Flexion) sind sinnvoll, wenn die Methode im Bereich des linguistischen Studiums eingesetzt wird. Dabei können die **Wortarten** der Bestandteile der Konstruktion angeführt werden, z. B. bei Bruttosozialprodukt (Substantiv) und auch Informationen zum **Flexionsparadigma** (Pl.: die Bruttosozialprodukte, Gen.: des Bruttosozialprodukt(e)s, etc.). Auch Details zur **Wortbildung** finden hier ihren Platz. Um die Zusammensetzung des Beispielwortes deutlich zu machen, könnte eine Zerlegung vorgenommen werden: Brutto + sozial + produkt.“* (KoGloss 2012: 27)

a) Google'i tõlgitud näide 4. (12.05.2017)

„Teave morfoloogia (teave sõnaliik, algustähega) on kasulikud kui meetodit kasutatakse valdkonnas keelelise uuringud. Kõne komponentide struktuuri võib viidata jaoks. Näiteks rahvuslikust koguproduktist (nimisõna) ja ka informatsioon flektiivsed (Pl.: RKT, geenid.: Edasi rahvuslikust kogutoodangust e) (s, jne). Samuti andmed Word moodustumise leida oma koht siin. Et koosseisu näiteks sõna selgelt eraldamine võiks teha: Gross + + toote sotsiaalne.“

b) Bingi tõlgitud näide 4. (10.05.2016)

„Morfoloogia (sõnaliik ja inflectional teave) andmed on vajalikud keelelise õppevaldkonnas kasutatakse meetodit. Sõnaliigid komponentide disain võib olla kasutatud, nagu rahvamajanduse kogutoodangu (nimisõna) ja ka teavet algustähega paradigma (PL.: GNP, Gen.: Bruttosozialprodukt (e) s, jne.). Samuti leiad sõna moodustumise üksikasju oma koht siin. Et selgitada sõna valimi koosseisu, lahuselu võeti: gross + sotsiaal + toode.“

Võrreldes kõige esimese toimetamata täieliku tekstilõiguga on selle tõlke kvaliteet tunduvalt kehvem. „(Die) Wortart“ on küll tõlgitud masintõlketööriista poolt õigesti „sõnaliigiks“ aga „(die) Flexion“ ei tähenda mitte käänupunkti vaid muutelõppu. „Bruttosozialprodukt“ on korrektselt tõlgitud rahvamajanduse kogutoodanguks (RKT),

kuigi täispikkuses sõnade asemel on kasutatud lühendit. Loogiline oleks, et lähtekeelsele lühendile antakse vastav sihtkeelne lühend ja lähtekeelsele täispikkuses sõnale vastav sihtkeelne täispikkuses sõna. Google'i tõlke jaoks oli „*des Bruttosozialprodukt(e)s*“ lõpus olev sulgudes „(e)“, kui mitmus (te) mitte genitiiv ehk omastav kääne. Saksa keeles asub sõna „*das Produkt*“ mitmuse vormi „*die Produkte*“ lõpus tõepoolest „e“.

Selle konkreetse lõigu puhul võiks väita, et Bingi ja Google'i tõlke kvaliteet on enamvähem samasugusel tasemel. Eelviimases lauses olev „*Bruttosozialprodukt*“ on jäetud tõlkimata. Kui sisestada see üksinda tuleb vasteks „Rahvamajanduse kogutoodang.“ Mõlemad masintõlketööriistad ei olnud suutelised tõlkima „*Flexionsparadigma*“ ehk muutevormistiku.²³ Selles näites olevad lõigud vajaksid umbes samasuges ulatuses järeltoimetamist.

Nende näidete juures hakkas silma, kui masintõlketööriistad ei ole suutelised oma tõlkemälust mingile terminile vastet leida, siis antakse ming „kokku pandud“ vaste. Sellised vasted on näiteks: „professionaalsete kasutajate“ (*Berufsnutzer*), „sõnastik kanne“ (*Glossareintrag*) ja „Word moodustumise“ (*Wortbildung*). Nendest oleks põhimõtteliselt võimalik tuletada õigeid vasteid. Google'i tõlgitud lõigu BLEU hindeks on 16.18 ja Bing on 19.57.

Mõlema masintõlketööriista üheks läbivaks iseärasuseks on see, et tõlgetesse lisatakse tühikud, kus algselt ei olnud. Kohati kaotatakse ära ka kirjavahemärke ja muid teksti elemente. Esmapilgul võivad sellised pisivead tunduda suhteliselt tühisea, kuid need hakkavad korduma terve teksti ulatuses, tekitavad need lisatööd.

Vastates küsimusele, millisest programmist oli tõlkimisel kõige rohkem kasu, siis selleks oli Google'i tõlge. Nimelt selgus analüüsi käigus, et Google'i tõlke poolt tõlgitud katkendid vajasisid väiksemas ulatuses järeltoimetamist. Google'i tõlgitud lauseid on sagedamini mõistetavamad. Bingi tõlgitud lasued olid kohati väga raskesti või üldse

23 Eesti keele seletav sõnaraamat: <http://www.eki.ee/dict/ekss/index.cgi?Q=paradigma&F=M>
(02.05.2017)

mitte mõistetavad. Google'i tõlgitud näidete BLEU hinnang on enamasti kõrgem kui Bingi oma. Seevastu Sam Goodman väidab oma artilis, et Bing olevat küsitletud inimeste arvates täpsem. (vrd Goodman 2014) See võib olla tõene, kui tõlgitakse inglise ja hispaania keele vahel. Hispaania keele puhul on tegu ühe maailmas enim kõneldava keelega ja seega on hispaaniakeelsete tekstide hulk tõlkemälus tunduvalt suurem kui eestikeelsete.

4.5. Üldised tõlkeprobleemid

Järgnevalt käsitletakse üldiseid tõlkeprobleeme, mis ei ole otseselt seotud masintõlkega. Saksakeelses algtekstis on kasutatud arvukaid jooniseid, kuid nummerdatud on ainult Moodle'it puudutava peatüki joonised. Peale selle on sulgudes tekstis nendele viidatud. Antud probleemi üle saaks vaielda, kas see kuulub tõlkeprobleemide või vormistamise alla. Sellele probleemile on võimalik läheneda mitmel viisil. Juhul kui tegu oleks tellimustööga, oleks mõttekas kliendiga selles osas konsulteerida, kui ta ei ole eelnevalt mingeid juhiseid andnud. Kui hakata koheselt neid jooniseid ise nummerdama võib see pärast põhjustada ebameeldivusi kliendiga. Üks võimalus oleks nummerdada kõik tekstis olevad joonised otsast peale (näiteks: joonis 1, joonis 2, ...). Teine võimalus oleks nummerdada joonised igas peatükis eraldi. Veel üks võimalus oleks jätta Moodle'i peatükis olevad joonised samaks ja nummerdada teised näiteks tähtedega (joonis A, joonis B, ...). Peale jooniste esineb tekstis ka paar tabelit, mis algtekstis ei ole samuti nummerdatud. Nende puhul võiks toimida sarnasel viisil nagu joonistega. Selles töös on joonised ja tabelid jäetud nummerdamata nii nagu algtekstis. Selle töö puhul ei ole tegu tellimusega ning mina leian, et selleks ei ole otsesest vajadust. Algteksti autorid ei leidnud samuti selleks vajadust.

Juhul kui klient soovib, et saksakeelsed joonised tõlgitaks samuti ära, siis oleks vajalik teha uued pildid vastavatest eestikeelsetest Moodle'i plokkidest ja vajadusel kohandada mingi pilditöötlusprogrammiga (nt Adobe Photoshop). Sellisel juhul peaks tõlkijal olema ligipääs Moodle'i keskkonnale ja tõenäoliselt oleks tarvis ka koolitaja või õpetaja õiguseid. Juhul kui tõlkijal ei ole ligipääsu ja temal ei ole mingitel põhujstel võimalik seda hankida, tuleb tal võtta ühendust isiku(te)ga, kellel on ligipääs ja vastavad õigused. Juhul kui tõlkijal tõepoolest ei ole mitte mingil viisil võimalik hankida juurdepääsu Moodle'i keskkonnale, siis alternatiiv oleks tõlkida Moodle'i plokid ära tuginedes internetis leiduvale eestikeelsele Moodle'it käsitlevale informatsioonile. Töö koostamise ajal annab Google'i pilditsing arvukaid tulemusi eestikeelsetest Moodle'i plokkidest. Peale nende on võimalik leida ka hulgaliselt erinevaid lehekülgi²⁴ ja PDF-faile²⁵. Minul

24 Kuidas alustada uue e-kursuse loomisega õpikeskkonnas Moodle:
http://voimalused.weebly.com/uploads/2/8/8/4/2884383/moodle_alustamine2012.pdf (25.06.2017)

endal on küll juurdepääs Moodle'i keskkonnale, kuid minul puuduvad paraku koolitaja või õpetaja õigused. Kasutasin ise seda meetodit teksti tõlkimiselt. Tõlkides ilma juurdepääsuta on oht, et tõlge ei ühti Moodle ametliku eestikeelse versiooniga. Muidugi üks võimalus oleks lisada tõlked originaali järgi sulgudesse, kuid need ei pruugi ühtida ametliku versiooniga, millele tõlkimise ajal puudus tõlkijal juurdepääs. Juhul kui oleks vaja tõlkida mingit teist veebikeskkonda, millel puudub eestikeelne versioon, käsitlevat teksti, siis jääks tõlkijal üle tugineda ainult oma enda teadmistele, kui mingit muud kohast materjali ei õnnestunud leida.

KoGlossi manuaalis kasutatavate jooniste puhul on tegu piltidega, mida ei ole võimalik sisestada mingisse masintõlketööriista nagu Google'i tõlge. Selleks, et nende tõlkimisel oleks võimalik kasutada masintõlget, tuleb kasutada mingit optilise märgituvastuse (*Optical character recognition*, OCR) programmi²⁶. Seejärel oleks võimalik tuvastatud tekst kopeerida ja kleepida mingisse masintõlketööriista. Seejuures tuleb vaadata, et tuvastatud tekst ei oleks vigane ning tuleb arvestada, et masintõlkest ei pruugi alati kasu olla. Tõlkeabiprogrammid nagu näiteks MemoQ²⁷ on suutelised tuvastama PDF-faile. Teine võimalus oleks kasutada rakendust, mis võimaldab tõlkida ka pilte. Tekstis olevate tabelite tõlkimise puhul mingit optilise märgituvastuse programmi ei ole tarvis. Tegin Wordis uued tabelid ja nende sisu tõlkimisel ei kasutanud mingit masintõlketööriista.

Üheks mõnevõrra väiksemaks tõlkeprobleemiks võiks olla *Antconci* ja teiste näitena toodud programmide tähistamiseks vaste leidmine. See oleks võibolla rohkem stiililine küsimus. Algtekstis on kasutatud kohati erinevaid mõisteid nagu näiteks „*Die Analyse-Software*“ ja „*Das Analyse-Programm*“. Alguses kaalusin „analüüsiprogrammi“ kasutamist, kuna see tundus mulle olevat kõige lähemal saksakeelsele „*Das Analyse-Programm*“. Lõpuks otsustasin KoGlossi koduleheküljel kasutatud „keeletarkvara“

25 <https://wiki.ut.ee/pages/viewpage.action?pageId=17113841>

26 Mõningad tasuta optilise märgituvastuse programmid:
Online OCR: <http://www.onlineocr.net/> (25.06.2017)
Free OCR: <http://www.free-ocr.com/> (25.06.2017)

27 Memoq: <https://www.memoq.com/> (25.06.2017)

kasuks. Otsustasin selle kasuks, sest soovisin oma tõlkega olla võimalikult kooskõlas seal oleva sõnakasutusega.

Järgnev näide pärineb alapeatükist 4.1. „Päringute esitamine (*Suchanfragen stellen*)“ punkt c) „Abi erandite puhul (*Hilfestellungen zu formalen Varianten*)“, kus selgitatakse *Wildcard*side ehk kohatäitjate kasutust päringute tegemisel *Antconc*is. Ühe näitena on seal toodud „*System, Systeme, Systems, Systemen*“, nende tõlgitud vormid oleksid vastavalt „süsteem, süsteemid, süsteemi, süsteemide.“ Juhul kui need näitena toodud sõnavormid ära tõlkida oleks vaja tõlkida ka nende juuresolevad näidisread, milles lähtekeelsed sõnavormid esinevad:

„[...] *koexistierten die Systeme friedlich* [...]“

Das regulative System der Versicherungswirtschaft [...];

Die Hauptkomponenten des Systems bilden [...].“ (KoGloss 2012: 17)

Neid näidisridasid on muidugi täiesti võimalik tõlkida eesti keelde, kuid seejuures tuleb arvestada, et need pärinevad KoGlossi saksa majanduskeele korpusest. Siinkohal võib tekkida mitu küsimust. Juhul kui näidisread ära tõlkida, kas lisada originaalid sulgudes ja/või joonealune märkus, selle kohta et tegu tõlgetega? Eelmiste näidisridade tõlkimisega erilisi probleeme põhimõtteliselt ei oleks, kuid järgmistega on olukord teistsugune. Selleks on nimelt näide, mis käib sisendi „++*such*++“ kohta. Selles näites on nagu eelmisteski esitatud korpusest leitud sõnavormid *suchen, sucht, suchte, suchten, gesucht*. Nende eestikeelsed vasted oleksid otsima, otsib, otsis, otsisid, oli otsinud. Google andis „++*such*++“ vasteks „uuris.“ Mina leian, et need saksakeelsed näidisread võivad jääda nii nagu nad on, sest nende eesmärk on illustreerida, millised on otsingu tulemused, eelnevalt selgitatud seadistusega. Sellisele tekstis leiduvale näitele nagu „*Die Erzeugung im Bauhauptgewerbe brach im März um 10,8 % ein*.“ (Handelsblatt online, 09.05.2005) oleks võimalik lisada tõlge selle järele või joone alla. KoGlossi meetodi kasutajad peaksid oskama rohkem kui ühte keelt. Samas on ju meetodi keskne keeletarkvara *Antconc* saadaval ainult ingliskeelsena. Lisaks seotakse KoGloss projekti

raames neli valminud konstruktsioonisõnastikku omavahel saksa keele kaudu. (vrd KoGlossi kodulehekülg 2012)

Moodle'i keskkonda käsitlevas peatükis andis Google'i tõlge lühendi LMS sulgudes oleva selgituse „(*Lernmanagementsystem*)“ vasteks ingliskeelse „(*Learning Management System*)“ Paraku ei ole olemas sobivat eestikeelset lühendit, mis koosneks tähtedest L, M ja S. Lähtekeelses selgituses on samuti sees ingliskeelne sõna „*Management*.“ Mina otsustasin oma tõlkes ingliskeelse vaste kasuks, sest inglise keel tundub olevat praeguse aja eestikeelsele lugejale tuttavam kui saksa keel. Minu enda kogemused Moodle'i keskkonnaga on olnud paraku valdavalt suhteliselt negatiivsed. Täpsemalt minu kogemused ei ole otseselt mitte Moodle'i keskkonnaga ise, vaid hoopis osade kursustega, mille raames seda keskkonda kasutati. Siinkohal võiks rääkida mingi teksti psühholoogilisest mõjust tõlkijale. Antud teksti tõlkimisel vähemalt minule see mingit tuntavat mõju ei avaldanud. Sellel teemal oleks võimalik kirjutada eralidi töö ja selle töö raames seda rohkem käsitleda ei saa.

4.6. Järeldused

Vastates küsimusele, millisest programmist oleks tõlkimisel kõige rohkem kasu, siis selleks on Google'i tõlge. Nimelt selgus analüüsi käigus, et Google'i tõlke poolt tõlgitud katkendid vajasisid mõnevõrra väiksemas ulatuses järeltoimetamist kui Bingi poolt tõlgitud. Google'i tõlgitud lauseid oli üldiselt mõistetavamad. Bingi tõlgitud lasued olid seevastu kohati väga raskesti või üldse mitte mõistetavad. Google'i kasutajate arv ületab Bingi oma ja seega on tõlkemälu mahukam, mis on statistilise mudeli puhul väga oluline. Seejuures tuleb silmas pidada seda, et need tulemused võivad sõltuvalt keelest varieeruda. Seda tingib asjaolu, et kõigile keeltele ei pakuta samal tasemel tuge. Tõlkekvaliteet võib mõne teisekeele puhul olla parem või kehvem. Masintõlketööriistade üheks läbivaks probleemiks on see, et kohati kaotatakse ära erinevaid kirjavahemärke ja mõnikord ka muid teksti elemente. Esmapilgul võib see tunduda suhteliselt tühise veana, kuid see võib tekitada segadust ja tekitada toimetajale lisa tööd. Siinkohal võivad osad inimesed loobuda masintõlkest ja eelistaksid võimalusel mõnda tõlkeabiprogrammi.

Vaidleksin vastu Mark Fišeli väitele, et „Masintõlge ei saa kunagi asendada inimese tõlget.“ (Fišel 2013: 1) See väide on igati tõene selle magistritöö kirjutamise ajal ja võibolla ka lähemas tulevikus saadaval oleva masintõlke kohta. Muidugi tuleb silmas pidada seda, et see artikkel on kirjutatud mitmeid aastaid tagasi. Praeguse aja ettevõtete jaoks on tähtis võimalikult suure kasumi teenimine ja kulude kokkuhoid. Võib üsna kindlalt väita, et selles niipea midagi ei muutu. Ettevõtetel on piisavalt motivatsiooni, loomaks mingi tehisintellekt, mis oleks suuteline tõlkima ilma inimeseta. Peale tõlkimise on arvutile leitud palju muid rakendusi. Näiteks on loodud programm, mis suudab värvida üleslaaditud joonistusi²⁸. Tulemused on enam-vähem võrreldavad masina poolt tõlgitud tekstiga. See oleks vaid üks asi paljudest, mida on ajaloo jooksul peetud võimatuks.

²⁸ PaintsChainer: <http://paintschainer.preferred.tech/> (25.06.2017)

Tänu masintõlkele on avanenud inimestele paljud võimalused mida enne ei olnud. Levib arvamus, et kuna „peamised infovood liiguvad maailmas teistes keeltes ja globaalset konteksti arvestades võib tunduda mõistlikum õppida inglise, saksa, hiina, hispaania või portugali keelt.“ (Koitmäe, Mändul 2013: 55) Siinkohal võib masintõlge olla potentsiaalselt kasulik väiksemate keelte püsijäämisel. Tänu sellele ei oleks enam vaja emakeelele eelistada võõrkeeli, sest kogu ülejäänud maailmast pärinevat informatsiooni oleks võimalik panna koheselt emakeelde. (vrd Koitmäe, Mändul 2013: 55) Käesoleva magistritöö koostamise ajal pakutakse juba teenuseid, mis osade keelte puhul võimaldavad tõlkida piltide ja keelte reaajas kõne. Kuid selle tagajärjel võib kaduda tõenäoliselt vajadus tõlkijate ja võibolla ka tõlkide järele. Samas võidaksid selles aga peale eesti keele ka muud väiksemad keeled.

Kokkuvõte

Käesoleva magistritöö eesmärgiks oli tõlkida saksakeelne KoGlossi manuaal eesti keelde ja analüüsida seda ning selgitada selle põhjal välja ka masintõlke otstarbekust selle laadse teksti tõlkimisel. Võrreldi ka kahe tuntud masintõlketööriista suutlikust omavahel valitud näidete põhjal. Tõlkimiseks valitud tekst tutvustab KoGlossi meetodit, mis peaks aitama parandada keeleõpet.

Teises peatükis käsitleti silmas peetud teoreetilis-metodoloogilisi lähtekohti. Kolmandas peatükis on esmalt tutvustatud erinevad masintõlke liigid. Teises osas vaadati lähemalt statistilist masintõlget. Neljandas ja viiendas osas tutvustati masintõlketööriistu Google *Translate* ja Bing *Translator*. Kuuendas osas käsitleti masintõlke otstarbekust peale tõlkija ka teistest vaatenurkadest. Peatüki lõpus on räägitud ka natukene omaenda kogemustest masintõlkega.

Neljas peatükk on tõlkeanalüüs, mis on jaotatud kuueks osaks. Esimeses osas tutvustati masintõlke hindamise meetodit BLEU. Teises osas tutvustati töö kesket terminoloogiat ja masintõlke võimekust selle tõlkimisel. Kolmandas osas uuriti kuivõrd võimekas on masintõlge KoGlossi manuaali tõlkimisel ja vaadati ka muid masintõlkega seonduvad probleeme, tähelepanekuid ja iseärasusi. Neljandas osas on toodud neli „suuremat“ näidet tõlke kvaliteedi kohta. Seejuures ei eeldatud, et masintõlketööriistad oleksid suutelised andma ideaalset tõlget ega arvestatud erialakeelsete sõnade õige tuvastamisega. Viiendas osas käsitleti teksti tõlkimisega tekkinud üldisi probleeme, mis ei ole seotud masintõlkega. Kuuendas osas on tehtud järeldused.

Masintõlge võib olla potentsiaalselt kasulik väiksemate keelte püsima jäämisele, sest tänu masintõlkele ei oleks vaja eelistada võõrkeelt emakeelele. Sellise masintõlke arengu ja leviku puhul ei saaks välistada olukorda, kus võib väheneda vajadus tõlkijate võibolla ka tõlkide järel, kuna juba töö kirjutamise ajal on osad rakendused suutelised tõlkima peale teksti ka pilte ja reaajas kõne. See aga ei pruugi tähendada seda, et vajadus nende järgi täielikult kaoks. Keerulisemad ja vastutusrikkamad tõlketööd jäätakse inimtõlkijate hoolde.

Analüüsimisel keskenduti eelkõige sellele, kas masintõlge tõepoolest aitab tõlkimisel aega ja vaeva kokku hoida või mitte. Selle analüüsi käigus selgus, et Google'i tõlke poolt tõlgitud katkendid vajasisid mõnevõrra väiksemas ulatuses järeltoimetamist kui Bingi poolt tõlgitud. Google'i tõlgitud laused andsid sagedamini rohkem aimu sisust. Bingi tõlgitud laused olid seevastu enamasti halvemini mõistetavad. Vastates küsimusele, kas masintõlge on abiks või mitte, siis ei ole veel võimalik anada ühest vastust. Nimelt on igal tõlkija omad eelistused, lähenemised ja harjumused. Palju sõltub ka tõlgitavast tekstist ja tõlke suunast. Tavaliselt tulevad kõik uuendused suurtematele keeltele ennem. Masintõlke poolt tõlgitud teksti toimetamine nõuab suurt tähelepanu, sest see võib eksida peale erialakeelsete sõnade ka tavapäraste sõnadega. See tähendaks lisatööd, mida igäüks ei oleks nõus tegema ja eelistatakse tõlkeabiprogrammi kasutamist. Tulemused võivad sõltuvalt keele paarist varieeruda, sest tõlgitavate keelte tugi ei ole ühel tasemel. Seejuures on oluline ka keelte omavahelised sidemed ja keele positsioon maailmas. Arvestada tuleb ka sellega, et masintõlge areneb pidevalt. Google ja Microsoft on hakanud vahetama statistilist mudelit neurovõrgu põhise vastu välja. Selle tulemusel peaks ka tõlkekvaliteet paranema, kuid käesoleva magistritöö koostamise ajal on eesti keele puhul kasutusel endiselt statistiline mudel.

Käesoleva magistritöö põhjal saaks väita, et selle koostamise ajal kättesaadavast masintõlkest ei ole KoGlossi manuaali laadse teksti tõlkimisel vähemalt saksa-eesti suunal „tuntavat“ kasu. Võib oletada, et mõlemate statistilisel mudelil põhineva masintõlketööriistade tõlkemälus leidub vähe selliseid eestikeelseid tekste, mis aitaksid genereerida paremat tõlget. KoGlossi manuaali puhul on tegu tekstiga, mis nõuab tõlkijalt tähelepanu ja erinevate nüanssidega arvestamist. Selle asemel, et aidata tekitaks masintõlge hoopis lisa tööd.

Kasutatud allikad

Esmased allikad

KoGlossi eestikeelne kodulehekülg. *Online* kättesaadav: https://www.uni-due.de/kogloss.eu/index_et.php (02.05.2017)

KoGlossi saksa keelne manuaal. *Online* kättesaadav: <http://dspace.ut.ee/handle/10062/28881> (02.05.2017)

KoGlossi inglise keelne manuaal. *Online* kättesaadav: <http://dspace.ut.ee/handle/10062/28890> (02.05.2017)

Teisesed allikad

Allan, Patrick 2016. Translation Tool Showdown: Google Translate vs. Microsoft Translator. Lifehacker. *Online* kättesaadav: <http://lifehacker.com/translation-tool-showdown-google-translate-vs-microso-1787836106> (10.04.2017)

Bramowicz, Matt 2014. Google Translate cannibalizes itself with poor machine translations. Lackuna – Language Translation Blog. *Online* kättesaadav: <http://www.lackuna.com/2014/02/10/google-translate-cannibalizes-itself-with-poor-machine-translations/> (02.05.2017)

Clark, Jack 2014. Google admits 'garbage in, garbage out' translation problem. *The Register*. *Online* kättesaadav: https://www.theregister.co.uk/2014/02/06/google_translate_issue/ (25.06.2017)

Fišel, Mark 2016. Milline on hea masintõlge? *Sirp*. *Online* kättesaadav: <http://www.sirp.ee/s1-artiklid/varia/milline-on-hea-masintolge/> (02.05.2017)

Goodman, Sam 2014. Machine Translation - Google Translate vs Bing Translator. *Online* kättesaadav: <http://blog.linnworks.com/google-bing-translate> (10.04.2017)

Haiying Li, Arthur C. Graesser; Zhiqiang Cai 2014. Comparison of Google Translation with Human Translation. *Online* kättesaadav:

<http://www.aaii.org/ocs/index.php/FLAIRS/FLAIRS14/paper/download/7864/7823>
(16.04.2017)

Kaalep, Heiki-Jaan; Koit, Mare 2010. Kuidas masin tõlgib. *Keel ja kirjandus* 10: 726-738. *Online* kättesaadav: <http://keeljakirjandus.eki.ee/726-738.pdf> (02.05.2017)

Kask, Katrin 2011. *Arusaamine tõlkeprotsessis ja seda mõjutavad tegurid*. Magistritöö. Tartu Ülikool, Tartu. *Online* kättesaadav: <http://dspace.ut.ee/handle/10062/17953> (25.06.2017)

Kibbermann, E.; Kirotar, S.; Koppel, P. 2007. *Saksa-eesti sõnaraamat. Deutsch-estnisches Wörterbuch*. Tallinn: Valgus.

Kirik, Harri 2008. *Juhendamata morfoloogia statistilises masintõlkes*. Bakalaureusetöö. Tartu Ülikool, Tartu. *Online* kättesaadav: http://kodu.ut.ee/~harts/web_files/Juhendamata_morfoloogia_statistilises_masintõlkes.pdf (25.06.2017)

Koitmäe, Toomas; Mändul, Merli 2013. Masintõlge – kas emakeele päästerõngas? *Oma Keel* 26: 52-55. *Online* kättesaadav: http://www.emakeeleselts.ee/omakeel/2013_1/OK_2013-1_07.pdf (02.05.2017)

Loik, Eve 2005. *Funktsionalismi põhimõtete kasutamine eesti kutseliste tõlkijate hulgas*. Magistripjekt. Tartu Ülikool, Tartu. *Online* kättesaadav: <http://dspace.ut.ee/handle/10062/17149> (25.06.2017)

Mäe, Helen 2014. *Vladimir vertlibi jutustuse „Mein erster Mörder“ tõlge ja tõlkeprobleemide analüüs*. Magistritöö. Tartu Ülikool, Tartu

Papineni, Kishore; Roukos, Salim; Ward, Todd; Zhu, Wei-Jing 2002. BLEU: *a method for automatic evaluation of machine translation* ACL-2002: 40th Annual meeting of the Association for Computational Linguistics. *Online* kättesaadav: <http://aclweb.org/anthology/P/P02/P02-1040.pdf> (19.06.2017)

Reiß, Katharina; Vermeer, Hans J. 1991. *Grundlegung einer allgemeinen Translationstheorie*. Tübingen: Niemeyer.

Saluäär, Hille 2007. Mis on tõlkepõhine tekstianalüüs ja mis kasu on sellest Eesti õigustõlkijal? *Õiguskeel* 4: 1-6. *Online* kättesaadav: http://www.just.ee/sites/www.just.ee/files/hille_saluaar._mis_on_tolkepohine_tekstianal_uus_ja_mis_kasu_on_sellest_eesti_oigustolkijal.pdf (12.05.2017)

Shankland, Stephen 2013. Google Translate now serves 200 million people daily. CNET. *Online* kättesaadav: <https://www.cnet.com/news/google-translate-now-serves-200-million-people-daily/> (02.05.2017)

Sibold, Gregor 2016. Google'i uus tõlkesüsteem on sama osav, kui inimene. *Geenius*. *Online* kättesaadav: <https://geenius.ee/rubriik/pilveteenused/googlei-uus-tolkesusteem-on-sama-osav-kui-inimene/> (02.05.2017)

V. Le, Quoc; Schuster, Mike 2016. A Neural Network for Machine Translation, at Production Scale. *Google Research Blog* *Online* kättesaadav: <https://research.googleblog.com/2016/09/a-neural-network-for-machine.html> (02.05.2017)

Vikipeedia: Google Translate. *Online* kättesaadav: https://et.wikipedia.org/wiki/Google_Translate (22.06.2016)

Vikipeedia: BLEU. *Online* kättesaadav: <https://en.wikipedia.org/wiki/BLEU> (22.06.2016)

Wilss, Wolfram 1988. *Kognition und Übersetzen: zu Theorie und Praxis der menschlichen und der maschinellen Übersetzung*. Tübingen: Niemeyer.

Resümee

Das Ziel der vorliegenden Masterarbeit ist es, das deutsche Manual der KoGloss-Methode ins estnische zu übersetzen und auf der Grundlage dieser Analyse die Nützlichkeit der heutigen maschinellen Übersetzung beim Übersetzungsprozess zu bestimmen. Der zu Übersetzung ausgewählte Text das Manual der KoGloss-Methode, soll helfen den Sprachunterricht zu verbessern.

Bei der Analyse konzentrierte man sich darauf, ob die maschinelle Übersetzung wirklich hilft einen Übersetzer Zeit und Mühe zu sparen oder nicht. Zu diesem Zweck wurden die zwei bekannten Online Übersetzungsmaschinen *Google Translate* und *Bing Translator*, die auf der statistischen Übersetzungsmethode beruhen, anhand von fünf ausgewählten Auszügen, miteinander verglichen. Dabei wurde nicht erwartet, dass die maschinelle Übersetzung in der Lage wäre, eine präzise Übersetzung zu liefern oder korrekt fachsprachliche Wörter zu erkennen. Die Übersetzungsmaschinen Google und Bing waren in der Lage die meisten Wörter in den Passagen zu übersetzen. Wie erwartet, gab es in der Übersetzung viele Deklinationsfehler und die Wortfolge war äußerst fehlerhaft. Die Analyse zeigte, dass die Übersetzungen von Google generell besser zu verstehen waren als die von Bing. In den Übersetzungen beider Übersetzungsmaschinen gab es auch eine Reihe von Besonderheiten und bestehenden Fehlern. Die Analyse zeigte auch, dass die übersetzten Abschnitte von Google generell weniger Bearbeitung erforderten als die Abschnitte von Bing. Die Übersetzungen von Bing waren manchmal sehr schwer oder gar nicht zu verstehen und brauchten sehr viele Korrekturen. Man kann sagen, dass die maschinelle Übersetzung tatsächlich hilfreich beim Übersetzungsprozess sein kann, aber nicht bei jeden Text.

Lõputöö autori kinnitus

Olen magistritöö kirjutanud iseseisvalt. Kõigile töös kasutatud teiste autorite töödele, põhimõttelistele seisukohtadele ning muudest allikaist pärinevatele andmetele on viidatud.

Autor: Eesnimi Perekonnanimi

(allkiri)

.....

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Martin Mala,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

**Manuaali „KoGloss: korpuspõhine kollaboratiivne konstruktsioonide glossarium
ülikoolide keeleõppes ja kutsetegevuses“ tõlge ja tõlkeanalüüs masintõlke
seisukohast,**

mille juhendaja on Terje Loogus,

- 1.1. reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
- 1.2. üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.
3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, 26.06.2017

Martin Mala